Segmentation based features for wide-baseline multi-view reconstruction

Armin Mustafa Hansung Kim Evren Imre Adrian Hilton
CVSSP, University of Surrey, Guildford, United Kingdom
a.mustafa@surrey.ac.uk

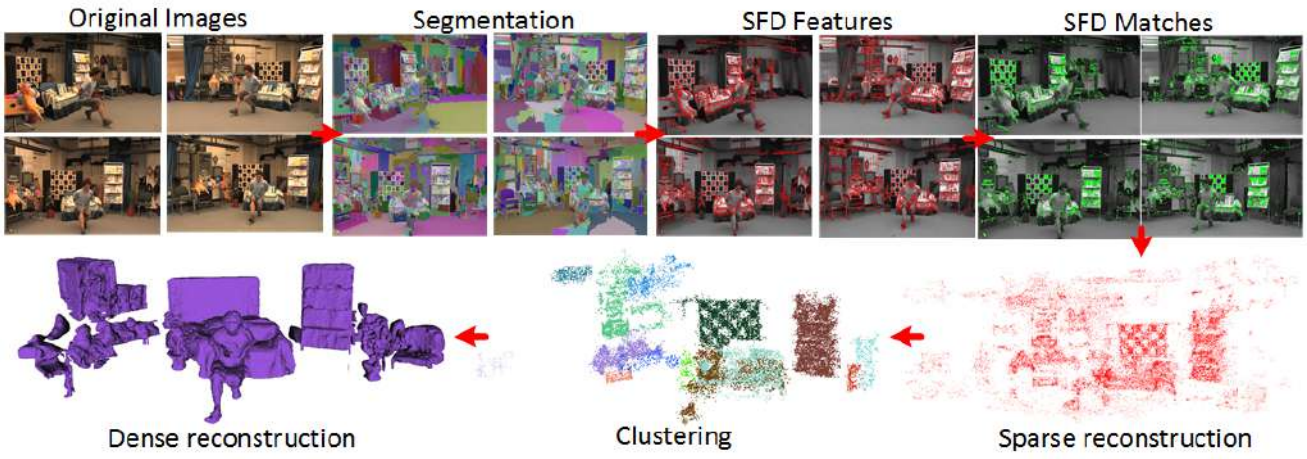

Figure 1: Segmentation-based Feature Detection SFD for wide-baseline matching and reconstruction for Odzemok.

Abstract

A common problem in wide-baseline stereo is the sparse and non-uniform distribution of correspondences when using conventional detectors such as SIFT, SURF, FAST and MSER. In this paper we introduce a novel segmentation based feature detector SFD that produces an increased number of 'good' features for accurate wide-baseline reconstruction. Each image is segmented into regions by over-segmentation and feature points are detected at the intersection of the boundaries for three or more regions. Segmentation-based feature detection locates features at local maxima giving a relatively large number of feature points which are consistently detected across wide-baseline views and accurately localised. A comprehensive comparative performance evaluation with previous feature detection approaches demonstrates that: SFD produces a large number of features with increased scene coverage; detected features are consistent across wide-baseline views for images of a variety of indoor and outdoor scenes; and the number of wide-baseline matches is increased by an order of magnitude compared to alternative detector-descriptor combinations. Sparse scene reconstruction from multiple wide-baseline stereo views using the SFD feature detector demonstrates at least a factor six increase in the number of reconstructed points with reduced error distribution compared to SIFT when evaluated against ground-truth and similar computational cost to SURF/FAST.

1. Introduction

Finding reliable correspondences between images is a fundamental problem in computer vision applications such as object recognition, camera tracking and automated $3D$ reconstruction. In this paper we focus on the problem of wide-baseline matching and reconstruction for general indoor and outdoor scenes. Established feature detectors such as Harris [16], SIFT [23], SURF [8], FAST [35] and MSER [25] often yield sparse and non-uniformly distributed feature sets for wide-baseline matching, as seen in Section 5.3 for SIFT and MSER. Gradient-based detectors (Harris, SIFT, SURF, STAR [3]) locate features at points of high-image gradient in multiple directions and scales to identify salient features which are suitable for affine-invariant matching across multiple scales resulting in very few features. Alternatively, Watershed segmentation based detectors (MSER) identify salient regions which are stable across multiple scales which can be reliably matched across wide-baseline views also resulting in relatively few features. Existing approaches result in a highly sparse non-uniform distribution of scene features. Whilst this may be sufficient for camera estimation and sparse point reconstruction using bundle-adjustment the resulting feature set often results in poor scene coverage as illustrated in Figure 2.

In this paper we propose a new segmentation based feature detector SFD which uses the segmentation boundary (local maximal ridge lines of the image) rather than the segmentation regions. SFD feature point detections are located

at the intersection points of three or more region boundaries. The intersection points represent local maxima of the image function in multiple directions giving stable localisation. Evaluation of SFD feature point detections across wide-baseline views demonstrates that although the segmentation changes with viewpoint the region intersection points are stable and accurately localised, an example is illustrated in Figure 1. SFD feature points are also demonstrated to give improved scene coverage with computational cost similar to existing efficient wide-baseline feature detectors (SURF/FAST). Contributions of this paper are:

- A novel segmentation based feature detector SFD for accurate wide-baseline matching and sparse reconstruction from wide-baseline views;
- SFD gives an increased number of repeatable feature detection for different viewpoints, accurate feature localisation and improved coverage for natural scenes;
- A comprehensive performance evaluation for wide-baseline matching on benchmark datasets of existing feature detectors (Harris, SIFT, SURF, FAST, MSER) and descriptors(SIFT, BRIEF, ORB, SURF) showing improved performance of the SFD detector in terms of both number of features and matching accuracy;
- Application to sparse scene reconstruction demonstrates an order of magnitude increase in the number of reconstructed points with improved scene coverage and reduced error compared to previous detectors against ground-truth.

2. Previous Work

Features are interesting image points with the properties described in [42]. A review of the decades of research into interest-point detection reveals three main approaches [30, 36]: image gradient analysis, intensity templates, and contour analysis.

Early image gradient-based approaches, such as Foerstner corner detector [13], define an optimal point based on the distances from the local gradient lines and Harris corner detector [16], define an interest-point as the maximum of a function of the Hessian of the image. A multi-scale extension was achieved by successive application of Gaussian kernels on scale-space representation of image and detecting interest-point as a local maximum both spatially, and across the scale-space [28]. Mikolajcyzk and Schmid seek these maxima via the Laplacian-of-Gaussian (LoG) filter, which is a combination of the Gaussian smoothing and the differentiation operation [27]. SIFT implements this as a difference-of-Gaussians [23]. A combination of gradient space with local symmetry was used in [18]. In [4], the scale-space representation is computed by nonlinear diffusion filtering (instead of Gaussian smoothing), yielding an improvement in the localisation accuracy. Gradient-based techniques offer accurate localisation [1], and are robust to many image transformations [28]. However, computation of the image gradients are sensitive to image noise and are computationally expensive. SURF mitigates this via the use of integral images and 2D Haar wavelets [8]. CenSurE achieves even faster operation by approximating the LoG operator with a bi-level filter [3]. However, A-Kaze claims superiority over all major gradient-based methods in terms of computational complexity, by using efficient diffusion filtering [5].

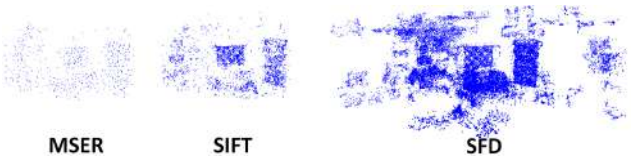

Figure 2: Sparse reconstruction comparison for Odzemok dataset.

Intensity template approaches seek patterns that are common manifestations of interest-points [36]. SUSAN [38] design a nonlinear cornerness function, which evaluates the dissimilarity of a pixel to a disc surrounding it. FAST replaces the nonlinear response functions by a simple, but effective heuristic: it first computes the intensity differences between the central pixel and a circle surrounding it, and then counts the contiguous pixels with a difference above a threshold [35]. A rotation-invariant implementation is proposed in [37], and a multi-scale extension, in [21]. MSER can be considered as a region detector responding to areas conforming to a "basin" template [25]. Intensity template methods are usually very fast, compared to their gradient-based counterparts [36]. However, with the exception of MSER, they are not affine-invariant, which limits their ability to cope with view-point variations. A recent evaluation indicates that their performance is relatively inferior to the alternatives [1].

Image contours give rise to two interest-point definitions: local maxima of the curvature along a contour, and intersections. Mokhtarian and Suomla [29] implement the former by building a scale-space representation of the contour map for the image, and detecting the local maxima of the curvature. The robustness was improved by using gradient correlation based detector [43]. Intersection of contour elements provides an alternative interest-point definition. T-junctions constitute a straightforward example [29][9] which inspires the proposed feature detector. Performance of curvature-based techniques are highly dependent on the quality of the extracted edges [6]. Although they are generally fast, the scale-space approach introduces a compromise between robustness and accuracy. On the other hand, contours, especially intersections are highly distinctive. Therefore, they are more robust to view-point variation [24, 6]. In this paper we proposed a segmentation based feature detector based on this property of intersections of contours which is robust to changes in viewpoint.

The number of features detected by curvature-based

techniques is quite small [6] and none of them have been evaluated on wide-baseline image pairs. They are based on only edge detection and vulnerable to the well-known difficulties in producing stable, connected, one-pixel wide contours/surfaces [15]. To avoid this we propose an over-segmentation based method for stable feature detection.

The idea of using regions for salient feature matching is well known and is exploited in [7, 41, 19] for applications other than wide-baseline stereo. A survey on interest points based on Watershed, Mean shift and Graph-cut segmentation was presented by [20]. A method is proposed [20] that uses boundaries and centres of gravity of the segments for extracting features. This demonstrates that Watershed is superior to the alternatives in terms of repeatability and Mean shift segmentation performs best for natural scenes. Watershed detects the local maxima of the gradient magnitude intensities as the region boundaries and proposed feature detection is based on the the detection of features as the intersection of local maxima, therefore we choose Watershed as our base segmentation technique.

3. SFD-Segmentation based Feature Detector

In this section we describe a new segmentation based feature detector. The main motivation for this approach is to increase the quantity and distribution of distinct features detected throughout the scene which are suitable for accurate wide-baseline matching and reconstruction. The approach is based on over-segmentation of the image into regions which ensures that detected features are distributed across the entire image as the region boundaries are located along contours of local maxima in the image which are consistent with respect to viewpoint change [20]. The use of local maximal contours overcomes the common problem of setting arbitrary thresholds or scales for feature detection, which is the basis for the proposed feature detector. Locating features where multiple region boundaries (3 or more) intersect gives good localisation, therefore SFD achieves good localisation which is consistent with-respect-to changes in viewpoint, as illustrated in Figure 1.

3.1. Feature detection

Segmentation of an image results in a large number of small regions with uniform appearance. The region boundaries represent ridge lines corresponding to local maxima of the image function or maxima in gradient if the segmentation is performed on a gradient image. The boundary intersection points where three or more region boundaries meet are local maxima in the image function in multiple directions. Consequently these points are accurately localised, highly distinctive and stable under changes in viewpoint giving good features for matching across wide-baseline views. This observation forms the basis of our proposed region-based feature detector, resulting in an increased number of salient features which are suitable for matching across wide-baseline views. The intersection points of three or more region boundaries in the over-segmented image are detected as features. For each boundary point the 3×3 pixel neighbourhood is tested for the number of region labels. If three or more region labels are present the point is detected as a feature as illustrated in Figure 3. These points are detected for the whole image on the region boundary contours.

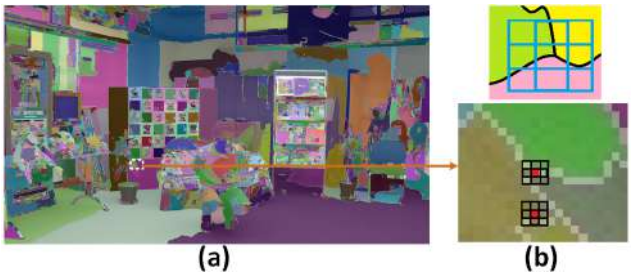

Figure 3: Intersection point: (a) Watershed segementation for Odzemok dataset; (b) Definition of SFD feature and examples.

3.2. Sub-pixel refinement

Let us denote the set of features detected for an image as $\mathscr{F} = \{f_1, f_2, ..., f_m\}$, where m is the total number of features. These features are integer values of the pixels where intersections of regions are detected. We perform a local sub-pixel refinement to optimise the feature location f_i at a local gradient maxima using the Levenberg-Marquardt [22] method. This refinement is based on the observation that every vector from the feature f_i to a point p_j located within a neighborhood $\mathscr{N}$ of $f_i = \{x, y\}^T$ is orthogonal to the image gradient $G_j = \{g_x, g_y\}^T$ at $p_j = \{x + \Delta x, y + \Delta y\}^T$, where $\Delta x, \Delta y$ is the shift at the point f_i. In our case we have chosen window size of 11×11 for the neighborhood $\mathscr{N}$ [14]. The cost function is defined as:

$$E(f_i) = \textstyle\sum_{j \epsilon \mathscr{N}} e_j(f_i), \text{ where,}$$
$$e_j(f_i) = (G_j^T (f_i - p_j)(1 - e^{-\frac{\Delta x_i^2 + \Delta y_i^2}{2}}))^2 \quad (1)$$

Since the vectors G_j and $f_i - p_j$ are orthogonal, $e_j(f_i)$ is 0 if f_i is at a local maxima, thereby making $E(f_i)$ to be 0. The sub-pixel position of the feature point is the minima of $E(f_i)$. The process is repeated for the entire feature set $\mathscr{F}$ to obtain a new solution $\mathscr{F}^*$ and the speed is optimized by parallelisation.

$$\mathscr{F}^* = \underset{f_i}{\operatorname{argmin}} \{E(f_i)\} \quad (2)$$

Feature descriptors are then applied to the local image regions of $\mathscr{F}^*$ to perform matching. In Section 5 we evaluate the detected feature points with descriptors based on SIFT and BRIEF for matching.

3.3. Segmentation

Segmentation of an image is defined as the process of partitioning an image into multiple segments. Pixels in each region share similar properties and are distinct from the pixels in adjacent regions. The boundary of the segments define contours of local maxima in the image. Our focus is on

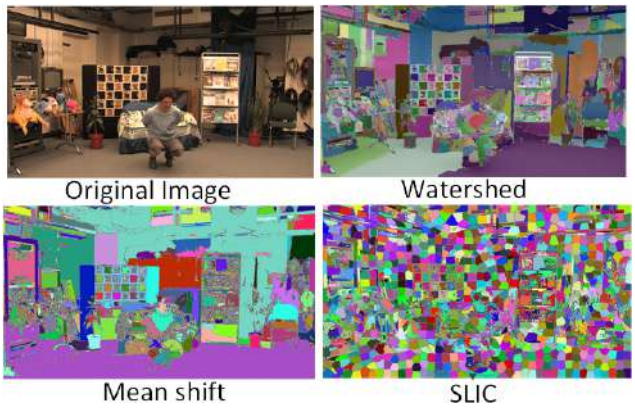

Figure 4: Different segmentation algorithms for SFD feature detection.

finding fast, automatic and stable over-segmentation techniques suitable for feature detection for general 3D scene reconstruction. The SFD features defined in Section 3.1 are evaluated on three different segmentation techniques:

Watershed (WA) [34]: The first segmentation technique is based on morphology. Readers are referred to [26] for detailed information on morphological segmentation techniques; we choose the watershed transform [34] because of speed and efficiency. The watershed transformation considers the gradient magnitude of an image as a topographic surface. Pixels having the highest gradient magnitude correspond to watershed lines which represent the region boundaries. Water placed on any pixel enclosed by a common watershed line flows downhill to a common local intensity minimum. Pixels draining to a common minimum form a basin, which represents a segment partitioning the image into two different sets: the catchment basins and the watershed lines.

Implementing the transformation on the image gradient, the catchment basins correspond to homogeneous grey level regions of this image. In practice, this transform produces an over-segmentation due to scene structure, local appearance variation and image noise. We use the modified version of the watershed algorithm defined in [32], replacing anisotropic diffusion with Bilateral filter [40]. An example is shown in Figure 4.

Mean shift (MS) [12]: This method is based on connectedness criterion and is proved to give stable and repeatable segments for natural scenes [20]. All pixels of an image are considered as vectors in 5D consisting of spatial and colour coordinates. Centroid based mode detection is employed and coordinates are ascribed modes. Recursive fusion of basins of attraction merges the modes located within a certain radius. This is an unsupervised segmentation technique and we perform over-segmentation on the image which is pre-processed using Bilateral filter to remove noise shown in Figure 4, followed by feature detection.

Simple Linear Iterative Clustering super-pixels (SLIC) [2]: This segmentation technique is based on Superpixel methods and it clusters pixels in the combined five-dimensional color and image plane space to efficiently generate compact, nearly uniform superpixels with a low computational overhead. SLIC is demonstrated to achieve good quality segmentation at a lower computational cost over state-of-the-art superpixel methods and to increase performance over pixel-based methods. The segmentation requires the number of regions (S) as input and in our case we calculate it using the following equation: $S = \frac{W*H}{w_{min}*h_{min}}$, where W and H are the width and height of input image and w_{min} and h_{min} are the minimum width and height of segmented regions which is set to approx 60×30 respectively to avoid very small segments as shown in Figure 4.

4. Wide-baseline Scene Reconstruction

Wide-baseline correspondences are obtained for all pairs of images using SFD. These correspondences are used to reconstruct a sparse 3D representation of the scene which is used for initialization of dense reconstruction. Figure 1 presents an overview of the algorithm for sparse to dense reconstruction.

4.1. Sparse Scene Reconstruction

We assume that the camera intrinsics are known and camera extrinsics together with 3D point locations are estimated using the correspondences. The fundamental matrix estimation procedure employs RANSAC and the normalised 8-point algorithm [17], to find the epipolar geometry using the intrinsics. The first camera is chosen as the world reference frame to obtain the camera matrix for the second camera from the fundamental matrix. Then, for each image correspondence, the triangulation algorithm [17] seeks the $3D$ point that minimises the re-projection error. After the initial pairwise sparse reconstruction is obtained, a new camera is registered to the structure [17] by finding the 2D and 3D correspondences between views and the 3D structure. The view with highest correspondences is selected and pose is estimated for the view from 3D-2D point correspondences using the RANSAC algorithm. The estimated pose minimizes reprojection error and the scene is augmented by triangulating the correspondences. The process is repeated for all the views until the camera pairs are exhausted. The algorithm employs global bundle-adjustment [39] to minimise the re-projection error over the calibration and the structure parameters to get the sparse reconstruction.

4.2. Dense Scene Reconstruction

The sparse reconstruction obtained above is used to initialize dense reconstruction of the scene. Sparse features are clustered in 3D space to obtain the initial coarse reconstruction. This is then refined for each object through joint optimisation of shape and segmentation using a robust cost function for wide-baseline matching. View-dependent opti-

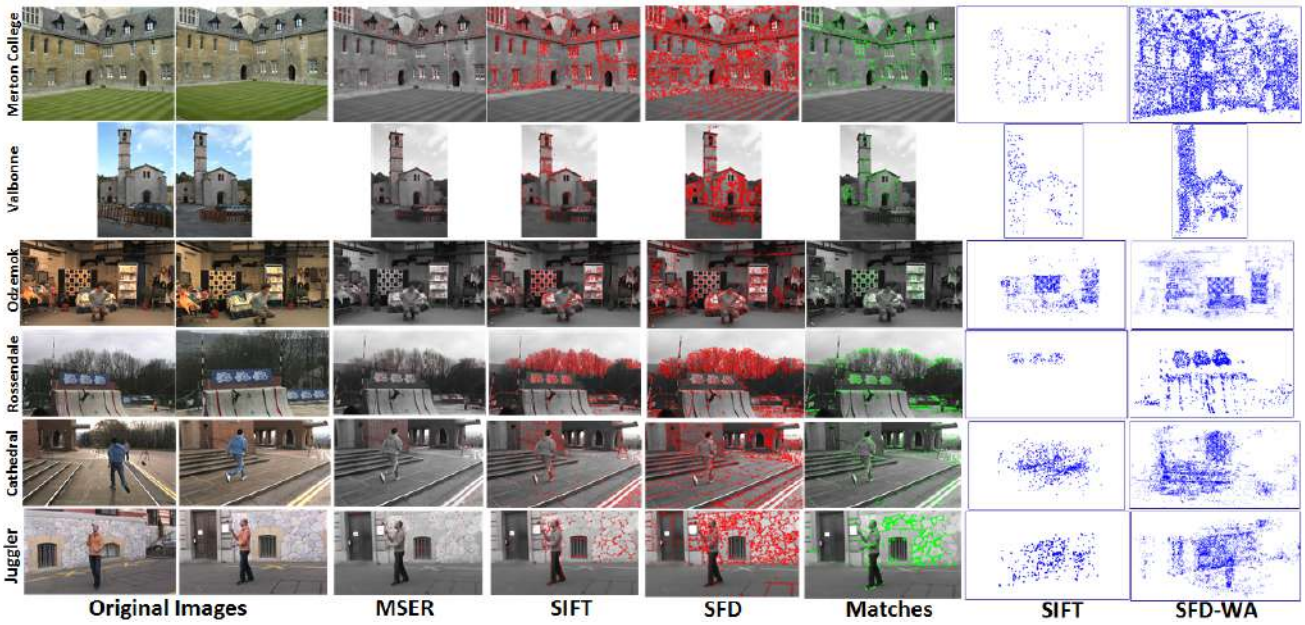

Figure 5: Results for all datasets: Column $1^{st} - 2^{nd}$ - Pair of images from each dataset, Column $3^{rd} - 5^{th}$ - Features detected on one image from each pair using WA, Column 6^{th} - Features matched between pair of images and Column $7^{th} - 8^{th}$ - Multi-view sparse reconstruction.

misation of depth is performed with respect to each camera which is robust to errors in camera calibration and initialisation to obtain dense reconstruction [31].

5. Experimental Results

The proposed algorithm is implemented in C++ using OpenCV [10] and tested on wide-baseline image/video datasets (15-30 degree angle between adjacent cameras) of natural indoor and outdoor scenes, as shown in Figure 5.

Merton College I[1]**, Valbonne**[1]: Outdoor scenes, repetitive background, varying lighting condition, static scenes.

Odzemok[2]: Indoor scene, scattered background, stable lighting condition, dynamic scene.

Rossendale[2]**, Cathedral**[2]**, Juggler**[3]: Dynamic outdoor scene, repetitive background, variation in illumination. Juggler is captured with only handheld cameras.

5.1. Evaluation criteria

We have evaluated our feature detector based on the properties of good features described in [42]: quantity; efficiency; accuracy; repeatability; and coverage of the proposed SFD against the state-of-the-art detectors (SIFT, SURF, MSER, Harris, STAR, ORB, FAST). The accuracy and repeatability is evaluated in Section 5.2 and quantity, coverage and efficiency is evaluated in Section 5.3. The coverage and quantity is further compared with SIFT by applying the feature detector to dense reconstruction in Section 5.4. Further results are presented in the supplementary material and video.

[1] http://www.robots.ox.ac.uk/ vgg/data/

[2] http://cvssp.org/data/cvssp3d/

[3] http://www.inf.ethz.ch/personal/lballan/datasets.html

5.2. Feature detection and matching accuracy test

Adjacent pairs of images are taken from each dataset and segmentation is performed using WA, MS and SLIC giving three variants SFD-WA, SFD-MS and SFD-SLIC respectively. The proposed SFD detection is performed on each pair of images for each segmentation method followed by feature matching using a SIFT descriptor. In order to evaluate the feature detector we use an exact nearest-neighbour matching algorithm followed by a ratio test as explained in [23]. All of the matches whose distance ratio is greater than 0.85 are rejected, which eliminates 90% of false matches and 5% of the correct matches [23]. After obtaining a set of refined matches, a left-right symmetry test is used to further remove inconsistent matches due to repeated patterns. This is followed by RANSAC based refinement [33] of matches without prior knowledge of camera parameters. The fundamental matrix is estimated using RANSAC and the inliers are chosen as the set of matches.

Experimental results for a pair of image for each dataset and all segmentation methods (WA, MS and SLIC) are summarized in Table 1. The column headed '$|F^*|$' shows the number of features detected in one of the images. Total count (TC) is the number of matches obtained with brute force matching using a SIFT descriptor and RANSAC count (RC) is the number of correspondences that are consistent with the RANSAC based refinement performed after the ratio and symmetry tests. The number of features detected by all segmentation techniques are similar. The numbers of matches reduces by $30 - 40\%$ after refinement using the symmetry and RANSAC tests (RC). The inlier ratio is

			WA			MS			SLIC		
Dataset	Resolution	#Cameras	$\|F^*\|$	TC	RC	$\|F^*\|$	TC	RC	$\|F^*\|$	TC	RC
Merton	1024×768	3(all static)	9947	7644	4533	8817	6899	4485	10336	8899	**5920**
Valbonne	512×768	7(all static)	3251	2915	1135	2939	2217	1252	4065	3352	**1981**
Odzemok	1920×1080	8(2 moving)	8169	6543	3717	7807	5908	3545	9905	7941	**4976**
Rossendale	1920×1080	8(all static)	7921	5057	2844	6878	4524	2698	7909	6629	**4066**
Cathedral	1920×1080	8(all static)	7207	6215	3370	6747	6050	3551	7983	6161	**3882**
Juggler	960×544	6(all moving)	4331	4325	2216	3996	3663	2155	5435	4393	**2763**

Table 1: SFD detection and matching results (best highlighted in bold): F^* - set of features, TC - Total count and RC - RANSAC count.

highest for SFD-SLIC segmentation compared to MS and WA.

Matching accuracy evaluation: For further evaluation of the number of feature matches obtained we use the known ground-truth reconstruction and camera calibration for the Odzemok dataset and SFD-WA as our base segmentation technique because of its computational efficiency. Ground-truth correspondences are obtained by back-projecting the $3D$ location of the feature points detected in one image to the other image and evaluating the distance to the estimated feature match. Mean re-projection error (MRE) given in Equation 3 is used for accuracy evaluation of the estimated SFD feature matches again the ground-truth.

$$MRE = \frac{\sum_0^N \sqrt{(x-x')^2 + (y-y')^2}}{K} \quad (3)$$

where (x, y) is the estimated SFD feature match, (x', y') is the re-projected point, and K is the number of feature matches, here $K = RC$. Table 2 presents the results of the ground-truth correspondence for the proposed SFD feature detector with a SIFT descriptor for matching and four other detector-descriptor combinations representing state-of-the-art detectors (MSER, FAST, SIFT). Matches (RC) shows the number of correspondences obtained with each approach after symmetry and RANSAC consistency tests. The number of matches obtained with the proposed SFD feature detector is greater by an order of magnitude than MSER and FAST, and by a factor three greater than SIFT. The MRE for SFD is lower compared to MSER and FAST and slightly higher than SIFT feature detector by approx. 0.2 pixels. The error for SFD is distributed between 0.3 to 2.8 pixels. Comparative evaluation of the re-projection errors for all the correspondences obtained by SFD and SIFT is shown in Figure 6. Figure 6 (a) shows that the number of wide-baseline matches for a given maximum re-projection error is consistent greater for SFD detection than for SIFT. Approximately 1000 points are concentrated at 1 pixel error depicting the relatively high accuracy of the proposed method as compared to SIFT. This implies that taking the best N features from SFD will give higher accuracy calibration/reconstruction than for SIFT feature detection. Therefore SFD gives more accurate geometry estimation from wide-baseline views due to the improved accuracy of feature localisation demonstrating the suitability of SFD for sparse 3D scene reconstruction.

Feature Detector	Descriptor	RC	MRE
MSER	SIFT	119	1.390
FAST	BRIEF	121	1.483
SIFT	SIFT	1269	1.175
SFD	SIFT	3717	1.351

Table 2: Ground-truth accuracy evaluation for feature matching on the Odzemok dataset.

Repeatability: We measure the repeatability (R) of SFD, defined as $R = \frac{\text{Correct Matches}}{\text{RC}}$ using the ground-truth information for Odzemok dataset. We eliminate the matches from RC with MRE greater that 2.5 pixels to obtain the 'Correct Matches', which is a standard setting to allow noise variance [17]. The comparisons with MSER, SIFT

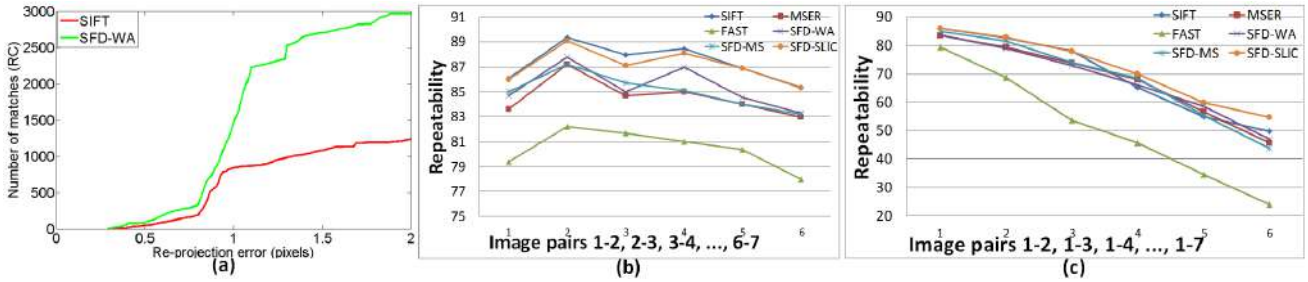

Figure 6: Accuracy and Repeatability results for Odzemok: (a) Re-projection error cumulative distribution of SIFT and SFD-WA; (b) Repeatability comparison for matching between adjacent views (15-30 degree baseline); and (c) Repeatability comparison for matching of camera 1 to all other views (15-120 degree baseline).

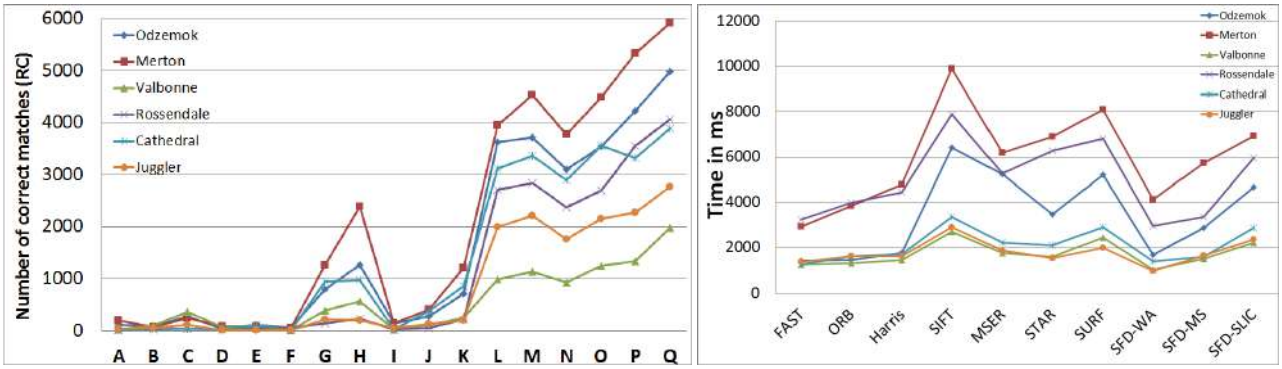

Figure 7: Evaluation on datasets (Left: Number of correct matches, A: FAST-BRIEF, B: Harris-BRIEF, C: Harris-SIFT, D: MSER-BRIEF, E: MSER-SIFT, F: ORB-ORB, G: FAST-ORB, H: SIFT-SIFT, I: STAR-BRIEF, J: SURF-BRIEF, K: SURF-SURF, L: SFD-WA-BRIEF, M: SFD-WA-SIFT, N: SFD-MS-BRIEF, O: SFD-MS-SIFT, P: SFD-SLIC-BRIEF and Q: SFD-SLIC-SIFT) and Right: Time for detecting features on a wide-baseline stereo pair for each sequence in ms.

and FAST are shown in Figure 6 (b) for adjacent image pairs with baseline 15-30 degrees and (c) between testing images 1-2, 1-3, ..., 1-7 with baseline 15-120 degrees.

The repeatability of SIFT and SFD-SLIC detector was comparable and greater than other detectors like FAST and MSER. Second best was SFD-WA followed by MSER and SLIC-MS. SLIC segmentation performed consistently better than other segmentation methods. As the baseline between the image pairs increases the repeatability reduces for each feature detector. The drop in the repeatability is similar for SFD, SIFT and MSER. The FAST detector does not perform well for wide-baseline images.

Evaluation of SFD vs. Harris/Uniform Sampling: The proposed SFD feature detector results in an increased number of features against previous detectors designed for wide-baseline matching applications. Alternative approaches to increase the number and coverage of feature detections could be use of corner detectors such as Harris or uniform grid sampling. Evaluation of the performance of SFD vs. Harris/Uniform sampling is presented in Table 3. For this comparison the threshold for the Harris detector and resolution for uniform grid resolution are set to give a similar number of features to SFD. Uniform grid sampling is performed by locating features at points of maximum gradient magnitude with a 13×13 grid resolution. The SIFT descriptor is used for all feature matching. Results presented in Table 3 show that the proposed SFD approach significantly outperforms the Uniform and Harris feature detectors after similarity and RANSAC tests are applied. This shows that the SFD approach detects stable features across wide-baseline views.

5.3. Benchmark Evaluation of Detector-Descriptor

To evaluate the performance of the proposed segmentation based feature detection approach for wide-baseline matching we present a comprehensive comparison with existing state-of-the-art feature detector and descriptor combinations. Comparison is performed with binary (FAST [5], ORB [37]) and floating point (Harris [16], SIFT [23], SURF [8], STAR [3], MSER [25]) detectors. These detectors are combined with feature descriptors (BRIEF [11], ORB [37], SIFT [23], SURF [8]). Detectors and descriptors are used with default parameters. Figure 7 presents the evaluation results for each detector-descriptor combination for wide-baseline matching on all datasets. The left column presents the number of correct matches (RC) obtained after similarity and RANSAC tests.

Feature Detector	Descriptor	Features	RC
Uniform Sampling	SIFT	8284	33
Harris	SIFT	8158	145
SFD	SIFT	8169	3717

Table 3: Evaluation of feature matching performance of SFD vs. dense feature sampling.

Performance of the proposed SFD detector combined with WA, MS and SLIC segmentation techniques with BRIEF and SIFT descriptors is shown in bars labelled L - Q, respectively demonstrating that the approach consistently achieves a factor $3 - 10$ increase in the number of correct matches compared to previous detector-descriptor combinations. The right column of Figure 7 presents the average computation time/frame showing that the computational time is less than floating point detectors and similar to binary detectors. SFD-WA is the fastest detector compared to MS and SLIC, but number of correct matches are highest for SLIC. MS gives lower number of correct matches compared to both WA and SLIC. The evaluation shows a trade-off between the performance and the number of correct matches for various segmentation techniques.

Scene Coverage: The distribution of the features across the scene is shown in Figure 5 for different detectors: Proposed SFD with WA, SIFT, MSER. Both the quantity and distribution of features for SFD give improved scene coverage for all the datasets.

5.4. Application to Wide-baseline Reconstruction

Wide-baseline sparse scene reconstructions are presented for all the datasets in Figure 5. Reconstructions obtained using the proposed SFD-WA features are compared with those obtained using the SIFT detector, in both cases the SIFT descriptor is used for matching. As expected from the evaluation of wide-baseline matching presented above the number of reconstructed points is much higher with the proposed approach as shown in Table 4 with WA, MS and SLIC. From Figure 5 it can be observed that sparse wide-baseline reconstruction based on SFD-WA gives a significantly more complete representation of the scene (evaluation of the accuracy against ground-truth reconstruction for Odzemok was presented in Table 2).

Dense reconstruction using the sparse SFD-WA features for initialisation is performed for Odzemok and Juggler datasets compared to initialisation using sparse SIFT features. This shows the importance of a a large number of features to obtain a more complete reconstruction including dynamic objects as illustrated in Figure 1.

Dataset	SFD-MS	SFD-SLIC	SFD-WA	SIFT
Merton	8118	**10965**	9619	316
Valbonne	3369	**5121**	4084	261
Odzemok	9087	**14515**	12385	1884
Rossendale	1017	**3983**	2213	238
Cathedral	9733	**12895**	10840	960
Juggler	6501	**8102**	7211	409

Table 4: Sparse points for pair-wise reconstruction.

We initialize the reconstruction and segmentation refinement algorithm [31] using sparse reconstruction obtained from the proposed algorithm. The results are shown in Figure 8. The sparse features based on SFD has better coverage compared to SIFT for both Juggler and Odzemok dataset. A large number of uniformly distributed sparse features in the 3D reconstructions leads to better initialization for dense reconstruction which is seen in Figure 8. The number of objects obtained from the sparse features in the final mesh reconstruction of SFD is higher than SIFT. Hence, the SFD based dense reconstruction gives more complete coverage of scene compared to SIFT.

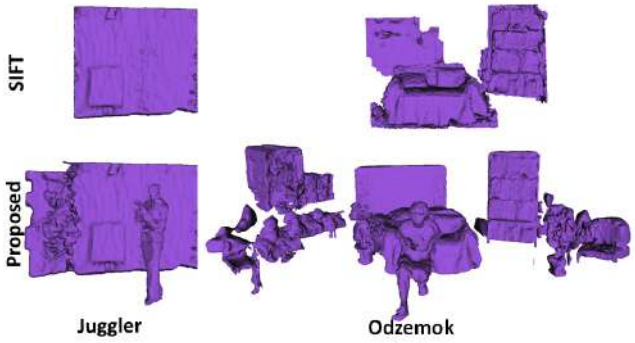

Figure 8: Dense reconstruction results for Odzemok dataset for SIFT and SFD feature detector

6. Limitations

Evaluation has been performed across a wide-variety of indoor and outdoor scenes to identify the limitations of SFD feature detection in the context of wide-baseline matching. As with other feature detection approaches the method is dependent on variation in surface appearance and consequently will produce fewer and less reliable features in areas of uniform appearance, or repetitive background texture like trees, sky etc. However, as demonstrated in the evaluation SFD increases the number of features and scene coverage for wide-baseline matching compared to previous feature detection approaches.

7. Conclusion and Future Work

In this paper we have proposed a novel feature detector for wide-baseline matching to support 3D scene reconstruction. The approach is based on over-segmentation of the scene and detecting features at intersections of three or more region boundaries. This approach is demonstrated to give stable feature detection across wide-baseline views with an increased number of features and more complete scene coverage than popular feature detectors used in wide-baseline applications. SFD is shown to give consistent performance for different segmentation approaches (Watershed, Mean shift, SLIC), with SFD-SLIC giving a marginally higher number of features. The speed of SFD feature detection is comparable to other methods for wide-baseline matching.

A comprehensive performance evaluation against previous feature detectors (Harris, SIFT, SURF, FAST, ORB, MSER) in combination with widely used feature descriptors (SIFT, BRIEF, ORB, SURF) demonstrates that the proposed segmentation based feature detector SFD achieves a factor $3-10$ times more wide-baseline feature matches for a variety of indoor and outdoor scenes. Quantitative evaluation of SFD vs. SIFT feature detection shows that for a given error level SFD gives a significantly larger number of features. Improved accuracy in feature localisation with SFD results in more accurate camera calibration and reconstruction of sparse scene geometry.

Application to stereo reconstruction from wide-baseline camera views demonstrates that the SFD feature detector combined with a SIFT descriptor achieves a significant increase in the number or reconstructed points and more complete scene coverage than SIFT detection. Further plans include evaluating the utility of SFD features in applications such as camera tracking and object recognition.

8. Acknowledgements

This research was supported by the European Commission, FP7 Intelligent Management Platform for Advanced Real-time Media Processes project (grant 316564).

References

[1] H. Aanæs, A. L. Dahl, and K. S. Pedersen. Interesting Interest Points - A Comparative Study of Interest Point Performance on a Unique Data Set. *IJCV*, 97:18–35, 2012. 2

[2] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *PAMI*, 34:2274–2282, 2012. 4

[3] M. Agrawal, K. Konolige, and M. Blas. Censure: Center surround extremas for realtime feature detection and matching. *ECCV*, 5305:102–115, 2008. 1, 2, 7

[4] P. F. Alcantarilla, A. Bartoli, and A. J. Davison. KAZE Features. In *ECCV*, pages 214–227, 2012. 2

[5] P. F. Alcantarilla, J. Nuevo, and A. Bartoli. Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces. In *BMVC*, 2013. 2, 7

[6] M. Awrangjeb, G. Lu, and C. S. Fraser. Performance Comparisons of Contour-Based Corner Detectors. *IEEE Trans. on Image Processing*, 21:4167–4179, 2012. 2, 3

[7] R. Basri and D. Jacobs. Recognition using region correspondences. *IJCV*, 25:8–13, 1995. 3

[8] H. Bay, T. Tuytelaars, and L. Gool. Surf: Speeded up robust features. In *ECCV*, pages 404–417. 2006. 1, 2, 7

[9] D. J. Beymer. Finding junctions using the image gradient. Technical report, 1991. 2

[10] G. Bradski. Opencv. *Dr. Dobb's Journal of Software Tools*, 2000. 5

[11] M. Calonder, V. Lepetit, C. Strecha, and P. Fua. Brief: Binary robust independent elementary features. In *ECCV*, pages 778–792, 2010. 7

[12] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *PAMI*, 24:603–619, 2002. 4

[13] M. A. Föstner and E. Gülch. A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centers of Circular Features. In *ISPRS Intercommission Workshop*, 1987. 2

[14] S. Gauglitz, T. Höllerer, and M. Turk. Evaluation of interest point detectors and feature descriptors for visual tracking. *IJCV*, 94:335–360, 2011. 3

[15] K. Haris, S. N. Efstratiadis, N. Maglaveras, and A. K. Katsaggelos. Hybrid image segmentation using watersheds and fast region merging. *IEEE Trans. on Image Processing*, 7:1684–1699, 1998. 3

[16] C. Harris and M. Stephens. A combined corner and edge detector. In *AVC*, pages 147–151, 1988. 1, 2, 7

[17] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2 edition, 2003. 4, 6

[18] D. C. Hauagge and N. Snavely. Image matching using local symmetry features. In *CVPR*, pages 206–213, 2012. 2

[19] J. Kim and K. Grauman. Boundary preserving dense local regions. In *CVPR*, pages 1153–1560, 2011. 3

[20] P. Koniusz and Mikolajczyk. Segmentation based interest points and evaluation of unsupervised image segmentation methods. In *BMVC*, 2009. 3, 4

[21] S. Leutenegger, M. Chli, and R. Y. Siegwart. Brisk: Binary robust invariant scalable keypoints. In *ICCV*, pages 2548–2555, 2011. 2

[22] K. Levenberg. A method for the solution of certain nonlinear problems in least squares. *Quarterly Journal of Applied Mathmatics*, 2:164–168, 1944. 3

[23] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60:91–110, 2004. 1, 2, 5, 7

[24] M. Maire, P. Arbelaez, C. Fowlkes, and J. Malik. Using contours to detect and localize junctions in natural images. In *CVPR*, pages 1–8, 2008. 2

[25] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *BMVC*, pages 36.1–36.10, 2002. 1, 2, 7

[26] F. Meyer. An overview of morphological segmentation. *International Journal of Pattern Recognition and Artificial Intelligence*, 15:1089–1118, 2001. 4

[27] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *IJCV*, 60:63–86, 2004. 2

[28] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool. A comparison of affine region detectors. *IJCV*, 65:43–72, 2005. 2

[29] F. Mokhtarian and R. Suomela. Robust Image Corner Detection Through Curvature Scale Space. *PAMI*, 20:1376–1381, 1998. 2

[30] H. Moravec. Obstacle avoidance and navigation in the real world by a seeing robot rover. Technical report, 1980. 2

[31] A. Mustafa, H. Kim, , J. Y. Guillemaut, and A. Hilton. General dynamic scene reconstruction from multiple view video. In *ICCV*, 2015. 5, 8

[32] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *PAMI*, 12, 1990. 4

[33] P. Pritchett and A. Zisserman. Wide baseline stereo matching. In *ICCV*, pages 754–760, 1998. 5

[34] J. B. Roerdink and A. Meijster. The watershed transform: Definitions, algorithms and parallelization strategies. *Fundam. Inf.*, 41:187–228, 2000. 4

[35] E. Rosten and T. Drummond. Fusing points and lines for high performance tracking. In *ICCV*, pages 1508–1511, 2005. 1, 2

[36] E. Rosten, R. Porter, and T. Drummond. Faster and better: A machine learning approach to corner detection. *PAMI*, 32:105–119, 2010. 2

[37] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. Orb: An efficient alternative to sift or surf. In *ICCV*, pages 2564–2571, 2011. 2, 7

[38] S. M. Smith and J. M. Brady. Susan—a new approach to low level image processing. *IJCV*, 23:45–78, 1997. 2

[39] N. Snavely, S. M. Seitz, and R. Szeliski. Modeling the world from internet photo collections. *IJCV*, 80:189–210, 2008. 4

[40] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *ICCV*, pages 839–846, 1998. 4

[41] E. Toshev, J. Shi, and K. Daniilidis. Image matching via saliency region correspondences. In *CVPR*, 2007. 3

[42] T. Tuytelaars and K. Mikolajczyk. Local invariant feature detectors: A survey. *Found. Trends. Comput. Graph. Vis.*, 3:177–280, 2008. 2, 5

[43] X. Zhang, H. Wang, A. W. B. Smith, L. Xu, B. C. Lovell, and D. Yang. Corner detection based on gradient correlation matrices of planar curves. *Pattern Recogn.*, 43:1207–1223, 2010. 2