Measuring the Coverage of Interest Point Detectors

Shoaib Ehsan, Nadia Kanwal, Adrian F. Clark and Klaus D. McDonald-Maier

School of Computer Science & Electronic Engineering,
University of Essex, Colchester CO4 3SQ UK
{sehsan, nkanwa, alien, kdm}@essex.ac.uk

Abstract. Repeatability is widely used as an indicator of the performance of an image feature detector but, although useful, it does not convey all the information that is required to describe performance. This paper explores the spatial distribution of interest points as an alternative indicator of performance, presenting a metric that is shown to concur with visual assessments. This metric is then extended to provide a measure of complementarity for pairs of detectors. Several state-of-the-art detectors are assessed, both individually and in combination. It is found that Scale Invariant Feature Operator (SFOP) is dominant, both when used alone and in combination with other detectors.

Keywords: Feature extraction, coverage, performance measure

1 Introduction

The last decade has seen significant interest in the development of low-level vision techniques that are able to detect, describe and match image features [1,2,3,4,5,6]. The most popular of these algorithms operate in a way that makes them reasonably independent of geometric and photometric changes between the images being matched. Indubitably, the Scale Invariant Feature Transform (SIFT) [1] has been the operator of choice since its inception and has provided the impetus for the development of other techniques such as Speeded-Up Robust Features (SURF) [2] and Scale Invariant Feature Operator (SFOP) [6].

One of the main driving factors in this area is the improvement of detector performance. Repeatability [7,8], the ability of a detector to identify the same image features in a sequence of images, is considered a key indicator of detector performance and is the most frequently-employed measure in the literature for evaluating the performance of feature detectors [5,8]. However, it has been emphasized that repeatability is not the only characteristic that guarantees performance in a particular vision application [5,9]; other attributes, such as efficiency and the density of detected features, are also important. It is desirable to be able to characterize the performance of a feature detector in several complementary ways rather than relying only on repeatability [5,10,11].

One property that is crucial for the success of any feature detector is the spatial distribution of detected features, known as the *coverage* [10]. Many vision applications, such as tracking and narrow-baseline stereo, require a reasonably even distribution of detected interest points across an image to yield accurate results. However, it is sometimes found that the features identified by detectors are concentrated on a prominent textured object, a small region of the image. Robustness to occlusion, accurate multi-view geometry estimation, accurate scene interpretation and better performance on blurred images are some of the important advantages of detectors whose features cover images well [10,11].

Despite its significance, there is no standard metric for measuring the coverage of feature detectors [10]. An approach based on the convex hull is employed in [12] to measure the spatial distribution for evaluating feature detectors. However, a convex hull traces the boundary of interest points without considering their density, resulting in an over-estimation of coverage. In [13], a completeness measure is presented but requires more investigation due to its dependence upon the entropy coding scheme and Gaussian image model used, and may provide varying results with other coding schemes for different feature types.

To fill this void, this paper presents a metric for measuring the spatial distribution of detector responses. It will be shown that the proposed measure is a reliable method for evaluating the performance of feature detectors. Since complementary feature detectors (*i.e.*, combining detectors that identify different types of feature) are becoming more popular for vision tasks [14,15,16], it is important to have measures of complementarity for multiple feature detectors, so that their combined performance can be predicted and measured [5]. This paper shows how *mutual coverage*, the coverage of a combination of interest points from multiple detectors, can be used to measure complementarity.

The rest of the paper is structured as follows: Section 2 describes the coverage measure, which is used to evaluate the performances of eleven state-of-the-art detectors on well-established data sets in Section 3. A complementarity measure derived from coverage, mutual coverage, is proposed in Section 4 and its effectiveness is demonstrated by results for combination of detectors. Finally, conclusions are presented in Section 5.

2 Measuring Coverage

There are several *desiderata* for a coverage measure:

- differences in coverage should be consistent with performance differences obtained by visual inspection;
- penalization of techniques that concentrate interest points in a small region; and
- avoidance of overestimation by taking into account the density of feature points.

The obvious way to estimate coverage is to calculate the mean Euclidean distance between feature points. However, different densities of feature points yield the same mean Euclidean distance. Conversely, the harmonic mean, which is widely used in data clustering algorithms [17], does penalize closely-spaced feature points, which augurs well for encapsulating their spatial distribution. Indeed, the harmonic mean is an inherently conservative approach for estimating the central tendency of a sample space, as:

$$A(x_1, \dots, x_n) \geq G(x_1, \dots, x_n) \geq H(x_1, \dots, x_n) \qquad \textbf{(1)}$$

where $A(.)$ is the arithmetic, $G(.)$ the geometric and $H(.)$ the harmonic mean of the sample set $x_1, \dots, x_n$, $x_i \geq 0\ \forall i$.
Formally, we assume that $p_1, \dots, p_N$ are the N interest points detected by a feature detector in image $I(x, y)$, where x and y are the spatial coordinates. Taking p_i as a reference interest point, the Euclidean distance d_{ij} between p_i and some other interest point p_j is

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \qquad \textbf{(2)}$$

providing $i \neq j$. Computation of (2) provides $N - 1$ Euclidean distances for each reference interest point p_i. The harmonic mean of d_{ij} is then calculated to obtain a mean distance D_i, $i = 1, \dots N$ with p_i as reference:

$$D_i = \frac{N-1}{\sum_{j=1, j\neq i}^{N} \left(\frac{1}{d_i}\right)} \quad \textbf{(3)}$$

Since the choice of the reference interest point can affect the calculated Euclidean distance, this process is repeated using each interest point as reference in turn, resulting in a set of distances D_i. Finally, the coverage of the feature detector is calculated as

$$\frac{N}{\sum_{i=1}^{N} \left(\frac{1}{D_i}\right)} \quad \textbf{(4)}$$

Since multi-scale feature detectors may provide image features at exactly the same physical location but different scales, interest points that result in zero Euclidean distance in (2) are excluded from these calculations on the basis that they do not provide independent evidence of an interest point.

In general, a large coverage value is desirable for a feature detector as a small value implies the concentration of interest points into a small region. However, the final coverage value obtained from (4) needs be considered against the dimensions of a specific image as the same coverage value may indicate good distribution for a small image but poor distribution for a large one.

3 Performance Evaluation

For the proposed coverage measure to have any value, its values need to be consistent with visual assessments of coverage across a range of feature detectors and a variety of images. To that end, this section presents a comparison of the coverage of eleven state-of-the-art feature detectors: SIFT (Difference-of-Gaussians), SURF (Fast Hessian), Harris-Laplace, Hessian-Laplace, Harris-Affine, Hessian-Affine, Edge-based Regions (EBR), Intensity-based Regions (IBR), Salient Regions, Maximally Stable Extremal Regions (MSER) and Scale Invariant Feature Operator (SFOP) [5,6]. Although different parameters of a feature detector can be varied to yield more interest points, it has a negative effect on repeatability and performance [13]. Therefore, authors' original binaries have been utilized, with parameters set to values recommended by them, and the results presented were obtained with the widely-used Oxford datasets [18]. The parameter settings and the datasets used make our results a direct complement to existing evaluations.

To demonstrate the effectiveness of this coverage measure, first consider the case of Leuven dataset [18] in Fig. 1. It is evident that SFOP outperforms the other detectors, where as values for EBR, Harris-Laplace and Harris-Affine indicate a poor spatial distribution of interest points. To back up these results, the actual distribution of detector responses for SFOP, IBR, Harris-Laplace and EBR for image 1 of the Leuven dataset are presented in Fig. 2. Visual inspection of these distributions is consistent with the coverage results of Fig. 1.

The coverage values obtained for Boat dataset [18] are presented in Fig. 3. Again, the performance of well-established techniques like SIFT and SURF is eclipsed by SFOP, a relatively new entrant in this domain. Other popular methods, such as Harris-Laplace, Harris-Affine, Hessian-Affine and EBR, again fare poorly. In addition, the curves depicted in Fig. 1 and 3 also exemplify the effects of illumination changes (Leuven) and zoom and rotation (Boat) on coverage.

A summary of the mean results obtained with all these feature detectors for the remaining datasets [18] is presented in Table I. It is clear that SFOP achieves much better coverage than the other feature detectors for almost all datasets under various geometric and photometric transformations.

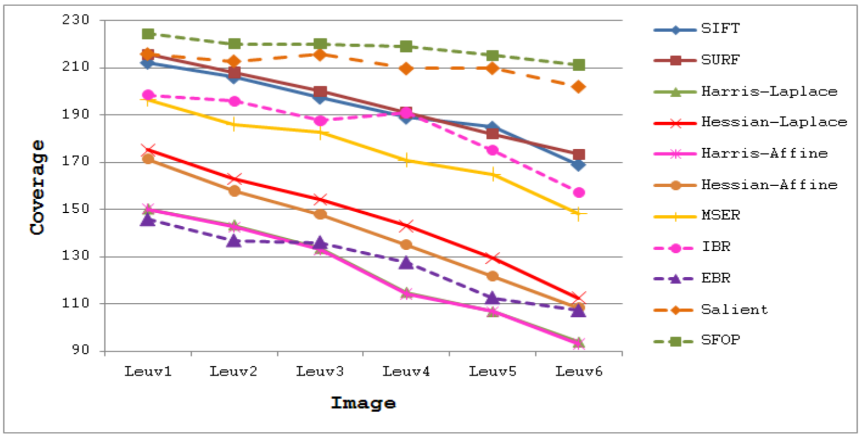

Fig. 1. Coverage results for Leuven dataset [18]

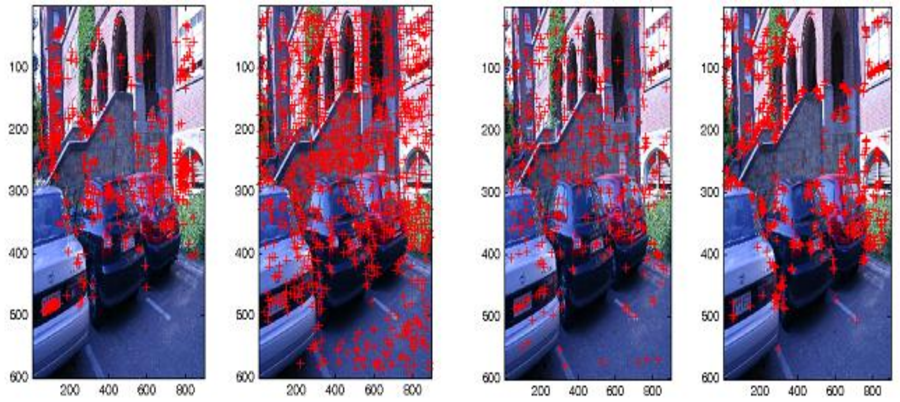

Fig. 2. Actual detector responses for image 1 of Leuven dataset [18]. From left to right: EBR, SFOP, IBR and Harris-Laplace

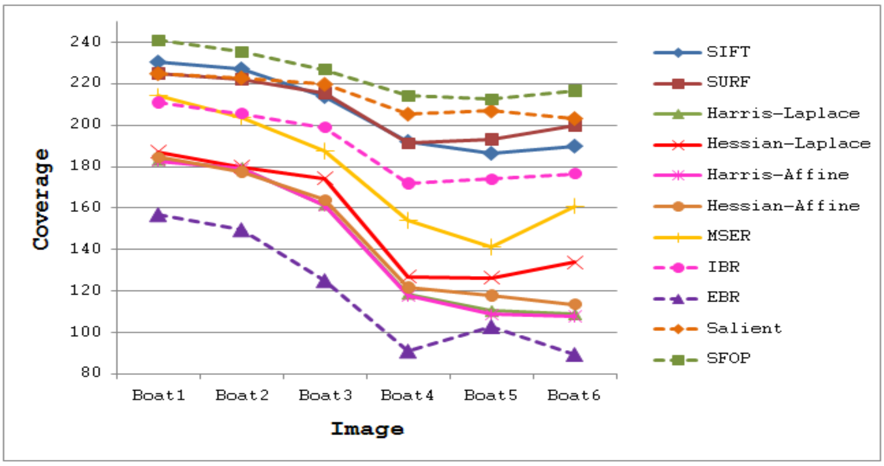

Fig. 3. Coverage results for the Boat dataset [18]

To exemplify the impact of these results on real-world applications, consider the task of homography estimation for the Leuven dataset. The mean error was computed between the positions of points projected from one image to the other, using a 'ground-truth' homography from [18], and a homography determined using the above detectors. SFOP performed the best, with a mean error of 0.245, where as EBR achieved a poor value of 3.672, consistent with the results shown in Fig. 1 and 2. In addition, we refer the reader to [11] that explains the significance of coverage of interest points (including those that cannot be matched accurately) for the task of scene interpretation. The proposed measure seems a viable method for determining coverage for such applications.

Table I Coverage results for state-of-the-art feature detectors

	Bark	**Bikes**	**Graffiti**	**Trees**	**UBC**	**Wall**
SIFT(DoG)	190.3	207.8	221.0	263.4	204.2	253.5
SURF(FH)	195.8	228.1	221.9	265.4	205.4	246.6
Harris-Lap	122.9	136.5	181.2	230.2	154.5	213.7
Hessian-Lap	120.0	154.5	199.2	234.2	154.9	208.6
Harris-Aff	122.8	136.0	181.0	229.9	153.8	212.8
Hessian-Aff	119.9	148.9	191.0	233.0	153.5	208.2
Salient Regions	190.6	258.7	218.0	256.4	201.5	236.4
EBR	139.2	138.3	166.4	214.3	119.0	204.4
IBR	192.3	214.7	209.7	255.5	198.4	243.8
MSER	179.6	86.4	200.3	229.6	200.6	248.3
SFOP	204.4	246.3	228.7	270.3	213.8	256.5

4 Mutual Coverage for Measuring Complementarity

Since the utilization of combinations of feature detectors is an emerging trend in local feature detection [5], this section proposes a new measure based on coverage to estimate how well these detectors complement one another. In addition to the principles mentioned in Section 2, the objective here is to penalize techniques that detect several interest points in a small region of an image. If detector A and detector B detect most feature points at same physical locations, they should have a low complementarity score. Conversely, a high score should be achieved if detector A and detector B detect most features at widely-spaced physical locations, indicating that they complement each other well. Again, a metric utilizing the harmonic mean seems a promising solution to achieve the required goal.

Formally, let us consider an image $I(x, y)$, where x and y are the spatial coordinates, being operated on by M feature detectors $F_1, F_2,\ldots, F_M$, so that $P_z = \{P_{z1}, P_{z2},\ldots P_{zN}\}$ is the set of N feature points detected by F_z. We then define

$$P_{zk} = P_z \cup P_k \quad \textbf{(5)}$$

as the set of feature points detected in image $I(x, y)$ by F_z and F_k. The coverage is then calculated as described in Section 2 using P_{zk}; as that includes points detected by both F_z and F_k, we denote it as the *mutual coverage* of F_z and F_k for image $I(x, y)$. Although this paper confines itself to combinations of two detectors, this notion of mutual coverage can be extended to more than two by simply combining their feature points in (5).

Mutual coverage has been applied to combinations of the detectors examined in the previous section. Inspired by [13], they can be categorized into four major classes, shown in Table II. For the purpose of this work, we confine ourselves to combinations of two detectors selected from two different categories; for example, SIFT is combined with EBR but not with SURF as they both detect blobs in a given image.

Fig. 4, 5 and 6 depict the average image coverage for SFOP, EBR and MSER when grouped with detectors from other categories for all 48 images of the Oxford datasets [18]. Interestingly, these results are consistent with the completeness results presented in [13]. Detectors from other categories perform well when combined with SFOP. The best results are achieved by grouping SFOP with a segmentation-based detector. A corner detector combined with a blob detector (except Hessian-Laplace and Hessian-Affine) yields good coverage. Segmentation-based detectors, however, do not seem to work well with corner detectors.

5 Conclusions

The performance of any image feature detector is dependent upon a number of different characteristics and one such property is coverage. This paper has proposed a coverage measure that produces results consistent with visual inspection. Furthermore, the mutual coverage of several feature detectors can be obtained simply by concatenating the feature points they detect and calculating the coverage of the combination. This gives us a rapid, principled way of determining whether combinations of interest point detectors will be complementary without having to undertake extensive evaluation studies; indeed, calculation is so rapid that one can consider using it online in an intelligent detector that adds features from other detectors in order to ensure that coverage, and hence accuracy of subsequent processing, is good enough.

An examination of the coverages of a range of state-of-the-art detectors identifies SFOP as the outstanding detector, both individually and when used in combination with other detectors.

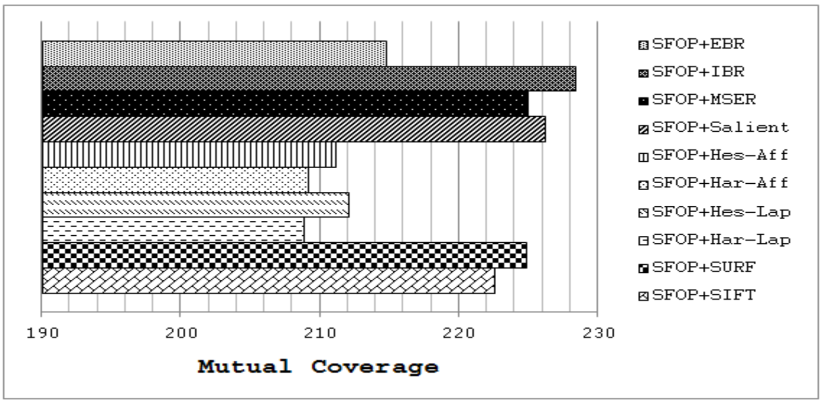

Fig. 4. Mutual coverage of SFOP in combination with other detectors

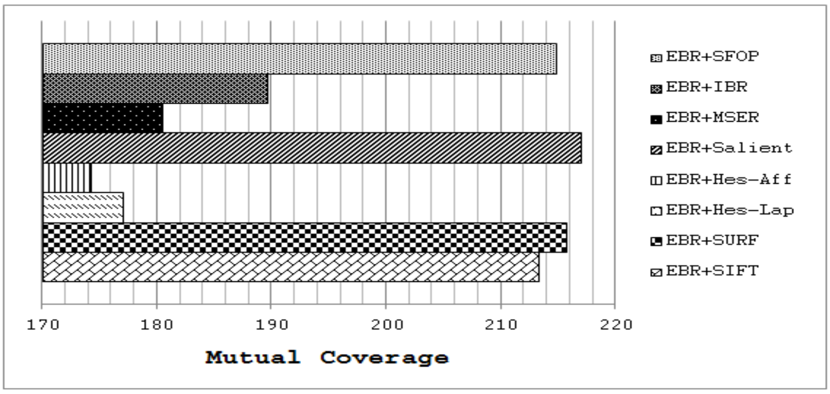

Fig. 5. Mutual coverage of EBR in combination with other detectors

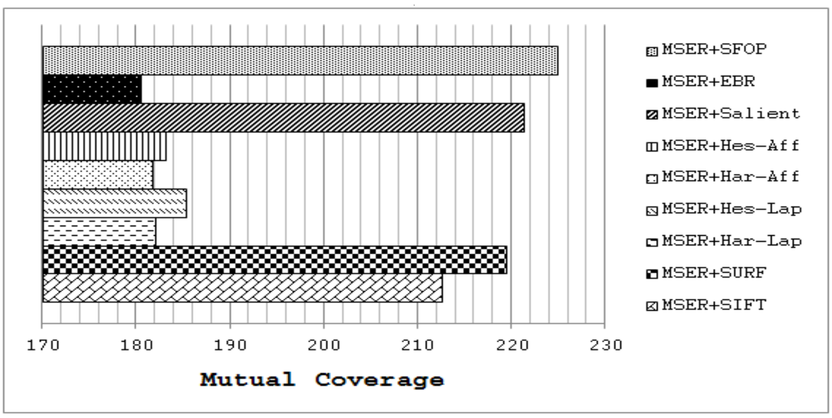

Fig. 6. Mutual coverage of MSER in combination with other detectors

Table II A taxonomy of state-of-the-art feature detectors

Category	Type	Detectors
1.	Blob detectors	SIFT, SURF, Hessian-Laplace, Hessian-Affine, Salient Regions
2.	Spiral detectors	Scale Invariant Feature Operator
3.	Corner detectors	EBR, Harris-Laplace, Harris-Affine
4.	Segmentation-based detectors	MSER, Intensity-based Regions

Acknowledgment. This work was supported in part by the UK EPSRC under grant EP/I500952/1.

References

1. Lowe, D.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, 60, 91-110 (2004)
2. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-Up Robust Features (SURF). Computer Vision and Image Understanding, 110, 346-359 (2008)
3. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust Wide Baseline Stereo from Maximally Stable Extremal Regions. BMVC, pp. 384-393, Cardiff, UK (2002)
4. Kadir, T., Zisserman, A., Brady, M.: An Affine Invariant Salient Region Detector. ECCV, LNCS, vol. 3021, pp. 228-241, Prague, Czech Republic (2004)
5. Tuytelaars, T., Mikolajczyk, K.: Local Invariant Feature Detectors: A Survey. Foundations and Trends in Computer Graphics and Vision, 3, 177-280 (2007)
6. Forstner, W., Dickscheid, T., Schindler, F.: Detecting Interpretable and Accurate Scale-Invariant Keypoints. ICCV, pp. 2256-2263, Kyoto, Japan (2009)
7. Schmid, C., Mohr, R., Bauckhage, C.: Evaluation of Interest Point Detectors. International Journal of Computer Vision, 37, 151-172 (2000)
8. Ehsan, S., Kanwal, N., Clark, A., McDonald-Maier, K.: Improved Repeatability Measures for Evaluating Performance of Feature Detectors. Electronics Letters, 46, 998-1000 (2010)
9. Nowak, E., Jurie, F., Triggs, B.: Sampling Strategies for Bag-of-Features Image Classification. ECCV, LNCS, vol. 3954, pp. 490-503, Graz, Austria (May 2006)
10. Perdoch, M., Matas, J., Obdrzalek, S.: Stable Affine Frames on Isophotes. ICCV, Rio deJaneiro, Brazil (2007)
11. Tuytelaars, T.: Dense Interest Points. CVPR, pp. 2281-2288, San Francisco, USA (2010)
12. Dickscheid, T., Förstner, W.: Evaluating the Suitability of Feature Detectors for Automatic Image Orientation Systems. ICVS 2009, LNCS, vol. 5815, pp. 305-314
13. Dickscheid, T., Schindler, F., Förstner, W.: Coding Images with Local Features. International Journal of Computer Vision (2010), doi: 10.1007/s11263-010-0340-z
14. Lazebnik, S., Schmid, C., Ponce, J.: Sparse Texture Representation using Affine-Invariant Neighborhoods. IEEE CVPR, pp. 319–324, Wisconsin, USA (June 2003)
15. Mikolajczyk, K., Leibe, B., Schiele, B.: Multiple Object Class Detection with a Generative Model. IEEE CVPR, pp. 26-36, New York, USA (2006)
16. Sivic, J., Zisserman, A.: Video Google: A Text Retrieval Approach to Object Matching in Videos. ICCV, vol. 2, pp. 1470-1477, Nice, France (2003)
17. Zhang, B., Hsu, M., Dayal, U.: K-Harmonic Means-A Spatial Clustering Algorithm with Boosting. LNCS, vol. 2007, 31-45 (2001)
18. Oxford Data Sets, http://www.robots.ox.ac.uk/~vgg/research/affine/

View publication stats