COMPARING FEATURE DETECTORS: A BIAS IN THE REPEATABILITY CRITERIA

Ives Rey-Otero[†], *Mauricio Delbracio*[†,§] *Jean-Michel Morel*[†]

[†]CMLA, ENS-Cachan, France
[§]ECE, Duke University, USA

ABSTRACT

Most computer vision application rely on algorithms finding local correspondences between different images. These algorithms detect and compare stable local invariant descriptors centered at scale-invariant keypoints. Because of the importance of the problem, new keypoint detectors and descriptors are constantly being proposed, each one claiming to perform better than the preceding ones. This raises the question of a fair comparison between very diverse methods. This evaluation has been mainly based on a *repeatability* criterion of the keypoints under a series of image perturbations (blur, illumination, noise, rotations, homotheties, homographies, etc). In this paper, we argue that the classic repeatability criterion is biased favoring algorithms producing redundant overlapped detections. We propose a sound variant of the criterion taking into account the descriptor overlap that seems to invalidate some of the community's claims of the last ten years.

Index Terms— Feature detectors, performance evaluation

1. INTRODUCTION

Local stable features are the cornerstone of many image processing and computer vision applications such as image registration [1, 2], camera calibration [3], image stitching [4], 3D reconstruction [5], object recognition [6–9] or visual tracking [10, 11]. The seminal SIFT method introduced by Lowe in 1999 [12, 13] sparked an explosion of local keypoints detector/descriptors seeking discrimination and invariance to a specific group of image transformations [14].

Ideally, one would like to detect keypoints that are stable to image noise, illumination changes, and geometric transforms such as scale changes, affinities, homographies, perspective changes, or non-rigid deformations. Complementarily, the detected points should be well distributed throughout the entire image to extract information from all image regions and from boundary features of all kinds (e.g., textures, corners, blobs). Hence, there is a variety of detectors/descriptors built on different principles and having different requirements. As opposed to interest point detectors, interest region detectors [15–18] extract the invariant salient regions of an image based on its topographic map. To fairly compare the very different feature detectors it is fundamental to have a rigorous evaluation protocol.

Introduced for the assessment of corner detectors [19] and later reformulated to evaluate scale/affine-invariant keypoint detectors [20–22], the repeatability criterion is the *de facto* standard procedure to assess keypoint detection performance [14].

The repeatability rate measures the detector's ability to identify the same features (i.e., *repeated* detections) despite variations in the viewing conditions. Defined as the ratio between the number of keypoints simultaneously present in all the images of the series (repeated keypoints) over the total number of detections, it can be seen as a measure of the detector's efficiency. Indeed, the repeatability rate incorporates two struggling quality criteria: the number of repeated detections (i.e., potential correspondences) should be maximized while the total number of detections should be minimized since the complexity of the matching grows with the square of the number of detections.

However, the repeatability criterion suffers from a systematic bias: it favors redundant and overlapped detections. This has serious consequences, as evenly distributed and independent detections are crucial in image matching applications [23]. The concentration of many keypoints in a few image regions is generally not helpful, no matter how robust and repeatable they may be.

In this communication, we unmask the repeatability bias, by explicitly considering the detectors redundancy in a modified repeatability measure. Experimental results show that the ranking of popular feature detectors is severely disrupted by the amended criterion.

The remainder of the article is organized as follows. Section 2 describes the repeatability criterion, discusses its variants, and illustrates how algorithms with redundant detections and unbalanced spatial distribution may perform better according to this traditional quality measure. Section 3 introduces a simple correction of the repeatability criterion that involves descriptor overlap. Section 4 reports on experimental evidence showing that the hierarchy of detectors is drastically altered by the new repeatability criterion. Section 5 contains a final discussion.

2. THE REPEATABILITY CRITERION AND ITS BIAS

2.1. Definition of the repeatability criterion

Consider a pair of images $u_a(\mathbf{x})$, $u_b(\mathbf{x})$ defined for $\mathbf{x} \in \Omega \subset \mathbb{R}^2$ and related by a planar homography H, that is, $u_b = u_a \circ H$. The detector repeatability rate for the pair (u_a, u_b) is defined as the ratio between the number of detections simultaneously present in both images, i.e., repeated detections, and the total number of detections in the region covered by both images.

In the repeatability framework, a detection generally consists of an elliptical region, denoted $R(\mathbf{x}, \Sigma)$, parametrized by its center $\mathbf{x}$ and a 2×2 positive-definite matrix Σ,

$$R(\mathbf{x}, \Sigma) = \left\{ \mathbf{x}' \in \Omega \mid (\mathbf{x}' - \mathbf{x})^T \Sigma^{-1} (\mathbf{x}' - \mathbf{x}) \leq 1 \right\}.$$

Acknowledgements: Work partly founded by Centre National d'Etudes Spatiales (CNES, MISS Project), European Research Council (advanced grant Twelve Labours), Office of Naval research (ONR grant N00014-14-1-0023), DGA Stéréo project, ANR-DGA (project ANR-12-ASTR-0035), FUI (project Plein Phare) and Institut Universitaire de France.

A pair of detections (elliptical regions $R(\mathbf{x}_a, \Sigma_a)$ and $R(\mathbf{x}_b, \Sigma_b)$) from images $u_a(\mathbf{x})$ and $u_b(\mathbf{x})$ will be considered repeated if

$$1 - \frac{|R(\mathbf{x}_a, \Sigma_a) \cap R(\mathbf{x}_{ba}, \Sigma_{ba})|}{|R(\mathbf{x}_a, \Sigma_a) \cup R(\mathbf{x}_{ba}, \Sigma_{ba})|} \leq \epsilon_{\text{overlap}}, \tag{1}$$

where $\mathbf{x}_{ba} = H\mathbf{x}_a$, $\Sigma_{ba} = A^{-1}\Sigma_b(A^T)^{-1}$ represents the reprojection of the ellipse on image u_b on the image u_a and A is the local affine approximation of the homography H.

The union and intersection of the detected regions are examined on the reference image $u_a(\mathbf{x})$ by projecting the detection on the image u_b into the image u_a. The union covers an area denoted by $|R(\mathbf{x}_a, \Sigma_a) \cup R(\mathbf{x}_{ba}, \Sigma_{ba})|$ while $|R(\mathbf{x}_a, \Sigma_a) \cap R(\mathbf{x}_{ba}, \Sigma_{ba})|$ denotes the area of their intersection. The parameter $\epsilon_{\text{overlap}}$ is the maximum overlap error tolerated. In most published benchmarks it is set to 0.40 [21, 22, 24].

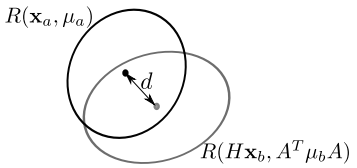

Fig. 1. Illustration of the repeatability criterion. Detection $R(\mathbf{x}_b, \Sigma_b)$ on image u_b is reprojected on the reference image u_a. If the overlap error is lower than $\epsilon_{\text{overlap}}$, the detections are considered repeated.

Since the number of repeated detections is upper bounded by the minimal number of detections, the repeatability rate is defined as

$$\text{rep} = \frac{\text{number of repeated detections}}{\min\left(|\mathcal{K}_a|_\Omega, |\mathcal{K}_b|_\Omega\right)} \tag{2}$$

where $|\mathcal{K}_a|_\Omega$ and $|\mathcal{K}_b|_\Omega$ denote the respective numbers of detections inside the area of Ω covered by both images u_a and u_b.

2.2. Illustration and alternative definitions

To discuss and illustrate the repeatability criterion and its variants, let us consider the particular case of a pair of detections $R(\mathbf{x}_a, \Sigma_a)$ and $R(\mathbf{x}_b, \Sigma_b)$ whose re-projections on the reference image are two disks, both of radius r and with centers separated by a distance d (Figure 1). Such a pair will be considered repeated if $d/r \leq f(\epsilon_{\text{overlap}})$, where f is a monotone function easily derived from (1). Figure 2 **(a)** shows the maximum distance d under which both detections will be considered repeated as a function of the radius r.

As pointed out in [22], detectors providing larger regions have a better chance of yielding good overlap scores, boosting as a result their repeatability scores. This also means that one can artificially increase the repeatability score of any detector by increasing the scale associated with its detections.

The authors of [22] proposed to avoid this objection by normalizing the detected region size before computing the overlapped error. The two detected elliptical regions $R(\mathbf{x}_a, \Sigma_a)$ and $R(\mathbf{x}_b, \Sigma_b)$ in (1) are replaced respectively by the elliptical regions $R(\mathbf{x}_a, {}^{\kappa^2}/_{r_a R_a}\Sigma_a)$ and $R(\mathbf{x}_b, {}^{\kappa^2}/_{r_b R_b}\Sigma_b)$, where r_a and R_a are the radii of the elliptical region $R(\mathbf{x}_a, \Sigma_a)$ and $\kappa = 30$ is its radii geometric mean after normalization.

The idea of such normalization was to prevent boosting a detector's performance by enlarging its associated ellipse. Yet, such a criterion is not scale-invariant, meaning that it may be over or under permissive depending on the detection size. For example, the maximal distance separating repeated detections of equal size does not take into account the scale (e.g., the radius of the circle in our special case illustration, see Figure 2 **(b)**). In consequence, with $\epsilon_{\text{overlap}}$ set to its standard value ($\epsilon_{\text{overlap}} = 40\%$), two circular detections of radius 1px and centers separated by 12px can still be regarded as repeated, although their respective descriptors may not even overlap!

Surprisingly, the code provided by the authors of [22][1] does not implement any of the criteria defined in their article. The code introduces a third definition by incorporating an additional criterion on the maximum distance separating two repeated keypoints that depends on the scale by

$$|\mathbf{x}_a - H\mathbf{x}_b| \leq 4\sqrt{r_a R_a}.$$

This criterion is illustrated in Figure 2 **(c)** for the same study case of two circular detections of equal size. This third criterion is not scale invariant either. Thus in this paper we shall stick to the first definition, which is scale invariant

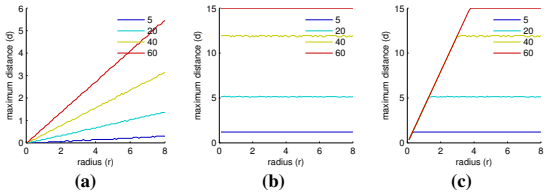

Fig. 2. Illustrating three different definitions of the repeatability criteria. Consider a pair of detections whose re-projections on the reference image are two disks of radius r with their centers separated by d. The maximal tolerated distance d_{max} between repeated detections is plotted as a function of the radius r for four values of the parameter $\epsilon_{\text{overlap}}$ (5%, 20%, 40% and 60%). **(a)** original definition given by (1), **(b)** with ellipses normalization $\kappa = 30$, **(c)** definition implemented in the provided code provided by the authors of [22]. Only the first definition is scale invariant.

2.3. Repeatability favors redundant detectors

The following mental experiment sheds light on how the repeatability favors redundancy. Let DET be a generic keypoint detector, and let DET2 be a variant in which each detection is simply counted twice. The number of repeatable keypoints and the total number of detections are both artificially doubled, leaving the repeatability rate unchanged. However, although the number of costly descriptor computations has doubled, no extra benefit can be extracted from the enlarged set of repeated keypoints. The classic repeatability rate fails to report that the benefit over cost ratio of DET2 is half the one of DET. This explains why methods producing correlated detections may misleadingly get better repeatability ratios.

A popular attempt for mitigating this drawback is to compare detectors at a fixed number of detections [24, 25]. This would not, however, solve the problem for two reasons. Firstly, by a similar reasoning as before, one can imagine a detector that repeats its best detection N times (N being the "fixed" number of detections) while discarding the rest. Such a detector would achieve optimal repeatability, despite being useless. But most importantly, given a detector, selecting the N best detections via a parameter (e.g., a threshold) is not generally an easy task. For example, in SIFT, the most popular

[1] Matlab code http://www.robots.ox.ac.uk/~vgg/research/affine/ retrieved date 2014/12/01

way of adjusting the number of detected keypoints is by thresholding the analysis operator (Difference of Gaussians) to retain only the most salient features. However, it is well known that this does not necessarily lead to a good selection in terms of stability [26]. To improve the selection, Li et al. [25] proposed a supervised regression process to learn how to rank SIFT keypoints. Although, this scheme produces good results it requires supervised learning.

For these reasons, we believe that a fair comparison should prefer the genuinely independent detections. The metric introduced in the following section is a first attempt in this direction.

3. NON-REDUNDANT REPEATABILITY

To demonstrate experimentally how taking into account redundancy drastically alters the hierarchy of detectors, we introduce in this section the non-redundant repeatability.

Besides the repeatability measure, which ignores the keypoints spatial distribution, other specific metrics have been proposed. Some examine the spatial distribution of the descriptors and others evaluate how well they describe the image. The ratio between the convex hull of the detected features and the total image surface is used in [27] as a coverage measure. The harmonic mean of the detections positions is used in [28, 29] as a measure of concentration. In [30], the authors propose to measure the completeness of the detected features, namely the ability to preserve the information contained in an image by the detected features. The *information content* metric proposed in [20] quantifies the distinctiveness of a detected feature with respect to the whole set of detections. Non specific features are indeed harmful, as they can match to other many and therefore confuse the matching. Being complementary to it, these metrics are generally used in combination with the repeatability rate. Nevertheless, since the purpose of the repeatability is to report on the benefit/cost ratio of a given detector, it should also, by itself, report on the description redundancy. We shall now see that the descriptors redundancy can be naturally incorporated in the repeatability criterion.

3.1. Non-redundant detected keypoints

To evaluate the redundancy of a set of detections $k \in \mathcal{K}$, each detection $(\mathbf{x}_k, \Sigma_k)$ can be assigned, in acccordance with the descriptor associated canonically with the keypoint for each method, a mask function $f_k(\mathbf{x})$ consisting of a truncated elliptical Gaussian

$$f_k(\mathbf{x}) = K e^{-\frac{1}{2\zeta^2}(\mathbf{x}-\mathbf{x}_k)^T \Sigma_k^{-1}(\mathbf{x}-\mathbf{x})}, \tag{3}$$

if $(\mathbf{x}-\mathbf{x}_k)^T \Sigma_k^{-1}(\mathbf{x}-\mathbf{x}) \leq \rho^2$ and 0 elsewhere. Each mask is normalized so that its integral over the image domain is equal to 1. The values ρ and ζ control the extent of the detected feature, as it can be derived from the descriptor's design. They will be fixed here for each detector by referring to the original paper where it was introduced. Indeed most detectors proposals come up with a descriptor or at least with a characterization of the region where this descriptor should be computed.

The sum of all descriptor masks $\sum_{k\in\mathcal{K}} f_k(\mathbf{x})$ yields a final map showing how much each image pixel contributes to the set of all computed descriptors. Note that one pixel may contribute to several descriptors. Similarly, the maximum taken over all detections $\max_{k\in\mathcal{K}} f_k(\mathbf{x})$ maps the pixels contribution to the best descriptor. Thanks to the mask normalization, the number of keypoints $K := card\,(\mathcal{K})$ is given by

$$K = \int_\Omega \left(\sum_{k\in\mathcal{K}} f_k(\mathbf{x}) \right) d\mathbf{x}, \tag{4}$$

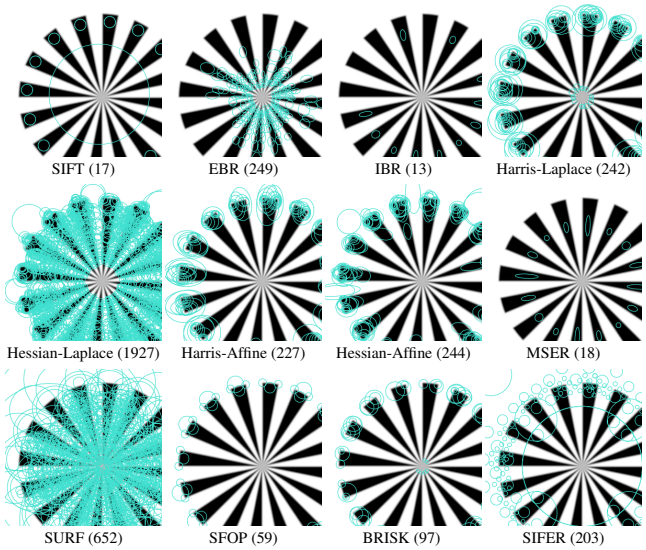

Fig. 3. Keypoints map on `siemens star` test image (total number of detections indicated in brackets). SIFT and SFOP seem to be the only (experimentally) rotationally invariant methods. The elliptical shapes deduced from the MSER regions have different sizes in each rotated triangle.

where Ω denotes the image domain. On the other hand,

$$K_{\text{nr}} := \int_\Omega \left(\max_{k\in\mathcal{K}} f_k(\mathbf{x}) \right) d\mathbf{x} \tag{5}$$

measures the number of *non-redundant* keypoints. This value can be interpreted as a count of the independent detections. To gain some intuition and see why this measurement is quite natural, let us examine four illustrative cases. Assume that there are only two detected keypoints so that $K = 2$. If the two detections

- completely overlap, then $K_{\text{nr}} = 1$.
- If they share the same center but have different sizes, then $1 < K_{\text{nr}} < K = 2$. But if their sizes are significantly different, then $K_{\text{nr}} \approx 2$, which makes sense. Indeed, one of them describes a fine detail and the other one a detail at a larger scale. Their information contents are roughly independent.
- If both keypoints are very close to each other then again $1 < K_{\text{nr}} < K = 2$ and the above remark on scales still applies.
- If the descriptors do not overlap at all then $K_{\text{nr}} = K = 2$.

3.2. Non-redundant repeatability

The above definitions entail a straightforward modification of the repeatability criterion (2). Let $\mathcal{K}_r$ be the set of repeatable keypoints between two snapshots, and Ω the area simultaneously covered by both images. We define the *non-redundant repeatability rate* by

$$\text{nr-rep} := \frac{\int_\Omega \max_{k\in\mathcal{K}_r} f_k(\mathbf{x}) dxdy}{\min\left(|\mathcal{K}_a|_\Omega, |\mathcal{K}_b|_\Omega\right)} \tag{6}$$

where $|\mathcal{K}_a|_\Omega$ and $|\mathcal{K}_b|_\Omega$ denote the respective numbers of detections inside Ω. The number of repeated detections in (2) is replaced in (6) by the number of non-redundant detections.

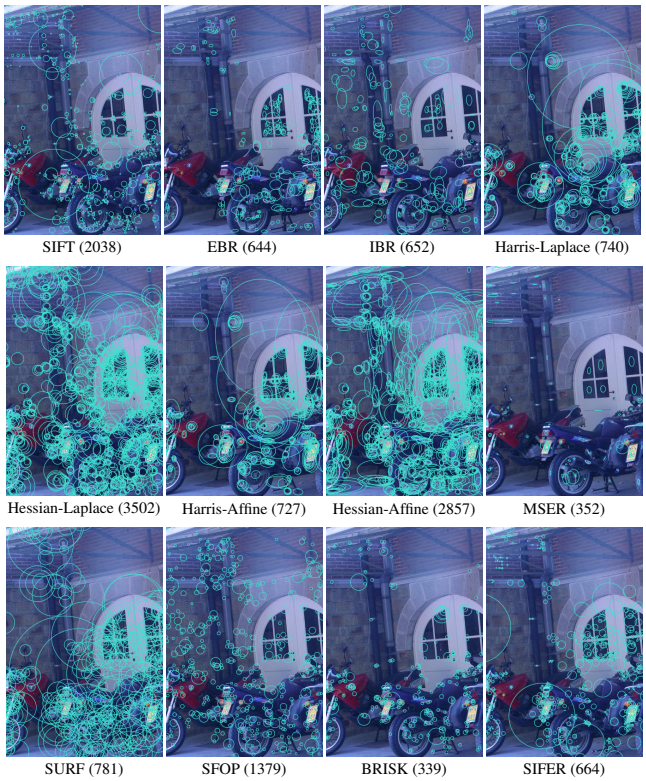

Fig. 4. Keypoints map on an image from the `bikes` sequence (total number of detections indicated in brackets). Most algorithms detect the same structure several times, producing significantly overlapped detections.

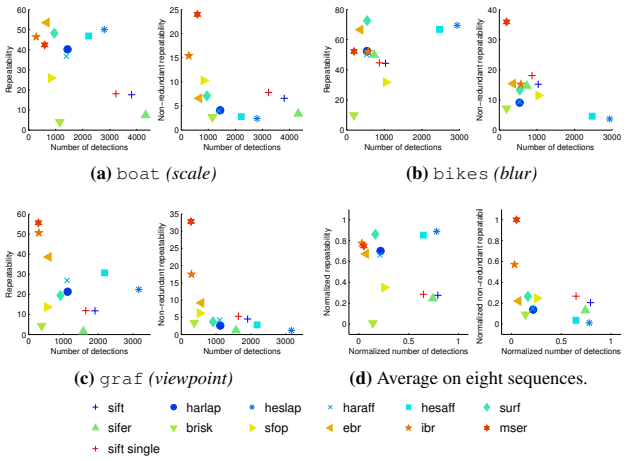

Fig. 5. **(a)**, **(b)**, **(c)** For three Oxford sequences, the average over the sequence of the repeatability and the non-redundant repeatability plotted as a function of the average number of keypoints detected. To compare a single detector performance the reader might follow the relative ordinate position of a particular detector in a particular scene in the traditional repeatability (left) and the non-redundant repeatability plots (right). For instance, MSER and SIFT algorithms always go up from the traditional to the non-redundant repeatability plots. This means that MSER and SIFT detections are less redundant than the average. **(d)** Average on eight sequences. For each sequence, the number of detections, the repeatability and the non-redundant repeatability are scaled to the full range of $[0, 1]$ and averaged into a single map.

4. EXPERIMENTS

To illustrate how the ranking is changed once redundancy is taken into account, twelve methods were examined, namely, the Harris-Laplace and Hessian-Laplace [22], Harris-Affine and Hessian-Affine [22], EBR [15], IBR [16], MSER [31], SIFT [12, 13], SURF [32], SFOP [33], BRISK [34] and SIFER [24][2]. We also considered a variant of SIFT that only takes one feature vector per detection. We called it SIFT-single.

Figures 3 and 4 show the detection maps for the twelve methods, on a synthetic image and a natural image from the Oxford dataset [22][3]. A simple visual inspection helps to get an overall idea of the algorithms redundancy.

Using the proposed non-redundant repeatability criterion, we evaluated the performance of the twelve feature detectors on the Oxford dataset. The performance evaluation of a detector is two-dimensional. On the one hand, a detector should produce as many detections as possible, while on the other, it should keep to a minimum the number of non-repeatable detections. In other words, the best detector is the one that has simultaneously the largest repeatability ratio and the largest number of detections.

Figure 5 **(a)**, **(b)**, **(c)** shows the average values, over three sequence, of the repeatability and the non-redundant repeatability plotted as a function of the average number of keypoints detected. Figure 5 **(d)** summarizes the detectors relative performance over the entire dataset (see caption for details).

On each of these plots, good detectors are on the top-right region. Thus, the benchmark reveals that the ranking of detectors is severely disrupted when considering the detectors redundancy. While for example Harris and Hessian based methods, SURF and EBR significantly reduce their performance (going down in the plots), MSER improves its relative position to the others. When the redundancy is not taken into account the method producing the most detections and the highest repeatability is the Hessian Laplace, while when considering the non-redundant variant it is SIFT.

5. DISCUSSION

In this paper, we have shown that the classic repeatability criterion is biased towards favoring algorithms producing redundant overlapped detections. This bias motivated the introduction of a variant of the repeatability rate taking into account the descriptors overlap. Experimental evidence showed that, once the descriptors overlap is taken into account, the traditional hierarchy of several popular methods is severely disrupted. Thus, the detections and associated descriptors generated by several methods are highly correlated. A reassuring characteristic of the new repeatability criterion is that it seems to be in agreement with the redundancies observed on patterns and on natural images. Future work will concentrate on evaluating the proposed criteria in more heterogeneous databases, also incorporating and discussing matching performance.

[2]Scripts and codes are available at `http://dev.ipol.im/~reyotero/comparing_20140906.tar.gz`

[3]Dataset available at `http://www.robots.ox.ac.uk/~vgg/research/affine/`

6. REFERENCES

[1] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, Cambridge University Press, 2003. 1

[2] N. Snavely, S.M. Seitz, and R. Szeliski, "Photo tourism: exploring photo collections in 3D," *ACM T. Graphics*, vol. 25, no. 3, pp. 835–846, 2006. 1

[3] R Grompone von Gioi, P. Monasse, J-M. Morel, and Z. Tang, "Towards high-precision lens distortion correction," in *ICIP*, 2010. 1

[4] G. Haro, A. Buades, and J-M. Morel, "Photographing paintings by image fusion," *SIAM J. Imaging Sci.*, vol. 5, no. 3, pp. 1055–1087, 2012. 1

[5] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S.M Seitz, and R. Szeliski, "Building Rome in a day," *Commun. ACM*, vol. 54, no. 10, pp. 105–112, 2011. 1

[6] W. Grimson and D. Huttenlocher, *Object recognition by computer: the role of geometric constraints*, MIT Press, 1990. 1

[7] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," in *CVPR*, 2003. 1

[8] H. Bay, B. Fasel, and L. Van Gool, "Interactive museum guide: Fast and robust recognition of museum objects," in *IWMV*, 2006. 1

[9] J. Zhang, M. Marszałek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: A comprehensive study," *Int. J. of Comput. Vision*, vol. 73, no. 2, pp. 213–238, 2007. 1

[10] D.B. Reid, "An algorithm for tracking multiple targets," *Trans. Autom. Control*, vol. 24, no. 6, pp. 843–854, 1979, . 1

[11] H. Zhou, Y. Yuan, and C. Shi, "Object tracking using SIFT features and mean shift," *Comp. Vis. Image Und.*, vol. 113, no. 3, pp. 345–352, 2009. 1

[12] D. Lowe, "Object recognition from local scale-invariant features," in *ICCV*, 1999. 1, 4

[13] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004. 1, 4

[14] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: A survey," *Found. Trends in Comp. Graphics and Vision*, vol. 3, no. 3, pp. 177–280, 2008. 1

[15] T. Tuytelaars and L. Van Gool, "Content-based image retrieval based on local affinely invariant regions," in *VISUAL*, 1999. 1, 4

[16] T. Tuytelaars and L. Van Gool, "Wide baseline stereo matching based on local, affinely invariant regions.," in *BMVC*, 2000. 1, 4

[17] T. Kadir, A. Zisserman, and M. Brady, "An affine invariant salient region detector," in *ECCV*. 2004. 1

[18] F. Cao, P. Musé, and F. Sur, "Extracting meaningful curves from images," *J. Math. Imaging Vision*, vol. 22, no. 2-3, pp. 159–181, 2005. 1

[19] R.M. Haralick and L.G. Shapiro, *Computer and robot vision*, Addison-Wesley Pub. Co., 1993. 1

[20] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors," *Int. J. Comput. Vision*, vol. 37, no. 2, pp. 151–172, 2000. 1, 3

[21] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *Int. J. Comput. Vision*, vol. 60, no. 1, pp. 63–86, 2004. 1, 2

[22] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," *Int. J. Comput. Vision*, vol. 65, no. 1-2, pp. 43–72, 2005, . 1, 2, 4

[23] M. Perdoch, J. Matas, and S. Obdrzálek, "Stable affine frames on isophotes," in *ICCV*, 2007. 1

[24] P. Mainali, G. Lafruit, Q. Yang, B. Geelen, L. Van Gool, and R. Lauwereins, "SIFER: Scale-Invariant Feature Detector with Error Resilience," *Int. J. Comput. Vision*, vol. 104, no. 2, pp. 172–197, 2013. 2, 4

[25] B. Li, R. Xiao, Z. Li, R. Cai, B.-L. Lu, and L. Zhang, "RankSIFT: Learning to rank repeatable local interest points," in *CVPR*, 2011. 2, 3

[26] I. Rey-Otero, J-M. Morel, and M. Delbracio, "An Analysis of scale-space sampling in SIFT," in *ICIP*, 2014. 3

[27] T. Dickscheid and W. Förstner, "Evaluating the suitability of feature detectors for automatic image orientation systems," in *CVS*. 2009. 3

[28] S. Ehsan, N. Kanwal, A.F. Clark, and K.D. McDonald-Maier, "Measuring the coverage of interest point detectors," in *IAR*. 2011. 3

[29] S. Ehsan, A.F. Clark, and K.D. McDonald-Maier, "Rapid online analysis of local feature detectors and their complementarity," *Sensors*, vol. 13, no. 8, pp. 10876–10907, 2013. 3

[30] T. Dickscheid, F. Schindler, and W. Förstner, "Coding images with local features," *Int. J. Comput. Vision*, vol. 94, no. 2, pp. 154–174, 2011. 3

[31] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image Vision Comp.*, vol. 22, no. 10, pp. 761–767, 2004. 4

[32] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded Up Robust Features," in *ECCV*, 2006. 4

[33] W. Förstner, T. Dickscheid, and F. Schindler, "Detecting interpretable and accurate scale-invariant keypoints," in *ICCV*, 2009. 4

[34] S. Leutenegger, M. Chli, and R.Y. Siegwart, "BRISK: Binary Robust Invariant Scalable Keypoints," in *ICCV*, 2011. 4