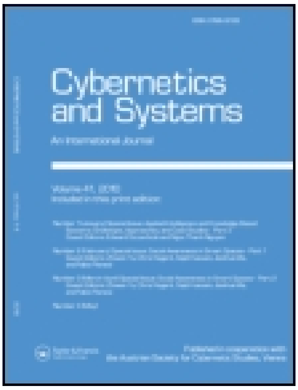
This article was downloaded by: [Iñigo Barandiaran]
On: 06 March 2013, At: 03:41
Publisher: Taylor & Francis
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Cybernetics and Systems: An International Journal

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ucbs20

AN EMPIRICAL EVALUATION OF INTEREST POINT DETECTORS

Iñigo Barandiaran [a] [b] , Manuel Graña [a] & Marcos Nieto [b]

[a] Grupo Inteligencia Computacional (GIC), Department of CCIA, UPV/EHU, 20018, San Sebastian, Spain

[b] Vicomtech-IK4, 20009, San Sebastian, Spain

To cite this article: Iñigo Barandiaran , Manuel Graña & Marcos Nieto (2013): AN EMPIRICAL EVALUATION OF INTEREST POINT DETECTORS, Cybernetics and Systems: An International Journal, 44:2-3, 98-117

To link to this article: http://dx.doi.org/10.1080/01969722.2013.762232

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Cybernetics and Systems: An International Journal, 44:98–117
Copyright © 2013 Taylor & Francis Group, LLC
ISSN: 0196-9722 print/1087-6553 online
DOI: 10.1080/01969722.2013.762232

Downloaded by [Iñigo Barandiaran] at 03:41 06 March 2013

An Empirical Evaluation of Interest Point Detectors

IÑIGO BARANDIARAN[1,2], MANUEL GRAÑA[1], and MARCOS NIETO[2]

[1]*Grupo Inteligencia Computacional (GIC), Department of CCIA, UPV/EHU, 20018, San Sebastian, Spain*

[2]*Vicomtech-IK4, 20009, San Sebastian, Spain*

Image interest point extraction and matching across images is a commonplace task in computer vision–based applications, across widely diverse domains, such as 3D reconstruction, augmented reality, or tracking. We present an empirical evaluation of state-of-the-art interest point detection algorithms measuring several parameters, such as efficiency, robustness to image domain geometric transformations—that is, similarity—affine or projective transformations, as well as invariance to photometric transformations such as light intensity or image noise.

KEYWORDS *computer vision, feature descriptors, interest points, point matching*

INTRODUCTION

Image analysis and computer vision–based applications deal with extraction of information from the images acquired by a camera sensor. Often, this information has a local representation in the form of selected pixels or set of pixels (regions) having relevant distinctiveness or discriminative characteristics. These relevant regions retain more information about the structures in the scene than surrounding neighboring regions.

In the literature, *feature points, key points*, or *features* are synonymous terms for the *image interest points*. In the last decade, there has been a lot of progress on interest point extraction algorithms used to build image and video descriptors leading to substantial advances in many computer vision

Address correspondence to Iñigo Barandiaran, Vicomtech-IK4, 20009, San Sebastian, Spain. E-mail: ibarandiaran@vicomtech.org

areas, which include 3D reconstruction, motion estimation, image registration, matching, and retrieval, or object and action recognition, needed at the first stages of many computer vision algorithm pipelines. Therefore, they can be considered as low-level image descriptors whose information is then delivered to other processes, in a bottom-up manner, until some kind of semantic knowledge or high-level interpretation is reached.

In this article we give the results of extensive empirical evaluation experiments measuring the behavior of several state-of-the-art interest point detection algorithms. This evaluation can be useful to help the research community improve their interest point extraction and feature descriptor approaches. In addition, practitioners of computer vision applications based on image point matching can obtain valuable information in this article to select the algorithm that best suits their needs.

The remainder of the article is organized as follows: Firstly, a general overview of interest point detectors is provided. Secondly, a description of each detection algorithm as well as the experimental framework and data sets used during the evaluation is provided. Thirdly, details on the results obtained during the evaluation are given. Finally, in the last section a general discussion about evaluated interest point detectors and some conclusions are provided.

METHODS

A recent exhaustive review of interest point detectors can be found in Tuytelaars and Mikolajczyk (2008), which describes and evaluates several affine region detectors. The authors identified the most relevant interest point detector performance measures:

- Repeatability: The same interest point should be extracted by the same detector even after geometric and/or photometric image transformations.
- Distinctiveness: Detected interest points should be different enough in order to be identified or matched, and the same interest point should not vary too much between different images of the same scene area.
- Quantity: Measures the number and distribution of interest points that a detector is able to extract from an image or set of images.
- Efficiency: Measures the computational cost in time and memory that an interest point detector needs to process an image.

It is worth noticing that the relative importance of these measures is application dependent; that is, every computer vision application or image analysis has different detection needs. For real-time optical tracking (Barandiaran et al. 2010) quantity and efficiency measures are critical, whereas for on-line object recognition, repeatability and distinctiveness are more relevant.

Downloaded by [Iñigo Barandiaran] at 03:41 06 March 2013

Downloaded by [Iñigo Barandiaran] at 03:41 06 March 2013

Nevertheless, the mentioned performance features can be reduced to two: quality and efficiency. Detector quality measures the ability of a detector to provide an accurate, precise, dense, and robust set of points to the next process in the image analysis pipeline. Detector efficiency measures how fast the interest point extraction is performed and which computational resources are required to carry out the task. Depending on these performance results, the ensuing processes in the pipeline should or should not need to apply further mechanisms such as filters, estimators, or heuristics in order for the application to succeed, or run efficiently and obtain accurate results.

Evaluation Data

In order for an interest point to be correctly identified from one image to another, the extraction algorithm must be transform invariant or covariant (Mikolajczyk et al. 2005); that is, robust to any type of image transformation categorized into two different classes: geometric and photometric transformations. In general, geometric transformations are those that modify the shape or the location of an interest point in the image space, usually generated by a change in the position and/or orientation of the objects in the scene or by a change in the point of view of the camera. On the other hand, photometric transformations influence the image feature appearance—that is, the intensity or color value of the pixels—due to changes in light conditions in the scene; the intrinsic parameters of the camera hardware, mainly the camera sensor; or a change in the acquisition parameters of the camera.

In the current detector evaluation we use several sets of images showing both geometric and photometric transformations. First, we use a data set proposed in Mikolajczyk et al. (2005) composed of three different sets of six images each, showing rotation, scaling, and perspective transformations as displayed in Figure 1.

FIGURE 1 Sample images from graffiti (left), boat (center), and brick (right) data sets from Mikolajczyk et al. (2005) (color figure available online).

In addition to these data sets, we use a collection of synthetic images generated by a tool proposed in Barandiaran et al. (2013). New transformed images are generated given an input image and a known geometric or photometric transformation between images. This tool allows us to generate new transformed images by applying similarities—that is, translations, rotations, or isotropic scaling transformations, as well as more general affine transformations.

Finally, the experimental data suite includes two data sets proposed in Barandiaran et al. (2013) containing photometric transformations.

Matching Evaluation

Image formation is modeled as in Eq. (1):

$$x_i = PX_w, \tag{1}$$

where X_w and x_i represent world points and their point projections in the image, respectively, and P represents the projection matrix, described in Eq. (2),

$$P = K[R|t], \tag{2}$$

where K describes the transformation from the camera reference frame to the image reference frame, and $[R|t]$ denotes the composition of a translation and a rotation transformation between world and camera coordinate systems.

The transformation between image points is given by a 2D linear projective transformation, aka *homography* (Hartley and Zisserman 2004), in the following situations: (a) world points X_w lie on a plane, so that the homography maps them into image points x_i, or (b) images are acquired by a camera rotating around its center of projection, so that the homography maps points x_i extracted from one image into points x_j extracted from another image of the same plane.

In the data set from Mikolajczyk et al. (2005) as well as in the one from Barandiaran et al. (2013), all images are related by a 2D homography H_{abD}. This a priori known transformation is used as the ground truth data, allowing to compute using Eq. (3) a priori predictions of where a point x_{iaD}, from image a of data set D, will be projected in image b of the same data set.

$$x_{jbD} = H_{abD}x_{iaD} \tag{3}$$

Similarly, points extracted from image b can be projected back to image a by application of the inverse of H_{abD}. Let $\tilde{x}_{jbD}$ be the estimated match of point x_{iaD} in image b obtained by the point detector algorithm. We use the knowledge of transformation H_{abD} to measure the accuracy and repeatability

Downloaded by [Iñigo Barandiaran] at 03:41 06 March 2013

of a point detector algorithm, computing the Euclidean distance d between the estimated and the ground truth points of a pair of images, as specified in Eq. (4):

$$d_{ij} = d\left(\tilde{x}_{jbD}, Hx_{iaD}\right)^2 + d\left(x_{iaD}, H^{-1}\tilde{x}_{jbD}\right)^2 \tag{4}$$

To identify the correct matches m_{ab} among all potential matches shown in Figure 2—that is, point pairs x_{ia} and $\tilde{x}_{jb}$ extracted from images a and b, respectively—we used the overlap error of Eq. (5) as proposed in Mikolajczyk and Schmid (2002).

This error measures the correspondence under the known homographic transformation of two supporting regions, usually ellipses or circles R_{ia} and R_{jb}, extracted around detected and projected interest points x_{ia} and $\tilde{x}_{jb}$, respectively. In our evaluation we consider a maximum 40% overlap error for a candidate pair of points to be considered as a potential true match.

$$\varepsilon_s = 1 - \left(\frac{R_{ia} \cap H^T R_{jb} H}{R_{ia} \cup H^T R_{jb} H}\right) \tag{5}$$

A pair of points x_{ia} and $\tilde{x}_{jb}$ whose Euclidean distance d_{ij} given by Eq. (4) and overlap error given by Eq. (5) are below set thresholds is considered a true match. We calculate the overlap between interest point neighbor ellipses by using software from Hughes and Chraibi (2011).

In the reported evaluation, we used estimated true matches to compute the repeatability score of a given interest point detector. The repeatability score for a given pair of images a and b (Mikolajczyk et al. 2005) was computed as the ratio between the number of point-to-point true correspondences and the minimum number of extracted points in the pair of images. Before computation of the repeatability score, we filtered out non-common points in both images, taking into account only parts of the scene present in

FIGURE 2 Correct and wrong matches between image a (left) and image b (right) (color figure available online).

Downloaded by [Iñigo Barandiaran] at 03:41 06 March 2013

both images. This filtering was applied using the a priori ground truth homography between both images.

Point Detectors

We have included in our evaluation most of the point detectors of the current state-of-the-art as well as a classical approach that today is still broadly used in the computer vision community. In the following section, a brief description of each point detector included in the evaluation is given.

- *HARRIS Detector*: The Harris corner detector is one of the feature extractors most commonly used by the computer vision community (Harris and Stephens 1988). Harris's approach improves Moravec's detector (Moravec 1977) by taking into consideration different orientations around the candidate pixel instead of shifting patches, by computing the second moment matrix, also called the *autocorrelation matrix*. Harris's cornerness measurement is still used by many point extractor approaches for non-maxima suppression. In this evaluation we have used a pyramidal version of Harris proposed in Mikolajczyk and Schmid (2001).
- *Scale Invariant Feature Transformation (SIFT) Detector*: This descriptor proposed by Lowe (1999) is one of the most successful approaches for interest point extraction and description to date. Detection is based on the convolution of images with a difference of Gaussian (DoG) operator $\eta = (g_\sigma - g_{\sigma'})$. Convolved images are arranged in a pyramidal representation, where each level (octave) of the pyramid is a downsampled and smoothed version of the image in the previous level. Smoothed images are obtained by convolving with a Gaussian operator with different values of scale σ. This arrangement of images allows SIFT to work in a scale-space representation (Lindeberg 1993). The SIFT detector and descriptor are designed to be invariant to rotation and scale transformations, but not to perspective transformation.
- *Speed Up Robust Feature (SURF) Detector*: This extractor (Bay et al. 2006) follows a similar approach to SIFT, explicitly addressing the problem of reducing computation cost. SURF searches for local maxima of the Hessian determinant in the scale space. SURF calculates Hessian determinants efficiently by using a discrete approximation of the Gaussian second-order partial derivatives, in conjunction with integral image representation (Viola and Jones 2004). Different from the SIFT approach, the scale estimation is not obtained by decreasing the image size after smoothing but by increasing the size of the discrete kernels.
- *FAST Detector*: This detector proposed in Rosten et al. (2010) follows a different approach than SIFT or SURF detectors. FAST uses supervised classification to label pixels as members of the class "interest point" or the class "background" by examining the values of pixels surrounding a candidate

Downloaded by [Iñigo Barandiaran] at 03:41 06 March 2013

point in a circular path, as illustrated in Figure 3. A feature is detected at pixel p if the intensities of at least n contiguous pixels of a surrounding circle of j pixels are all below or above the intensity of p by some threshold t. The original FAST approach does not perform scale-space representation.

- *BRISK Detector*: Proposed in Leutenegger et al. (2011), this detector implements a modification of the FAST detector proposed in Mair et al. (2010) that improves the original FAST score computation by changing the original classifier to a binary decision tree. The BRISK detector tries to overcome the limitations of the FAST detector regarding scale robustness by computing the FAST score over several octaves in a scale-space representation.
- *Maximally Stable Extremal Regions (MSER) Detector*: This detector proposed in Matas et al. (2002) is an approach based on the detection of blob-like structures. MSER detects blobs by using local luminance extrema, obtained by iteratively applying watershed-based segmentation. A region R_i is considered stable and, therefore, a potential feature if for all its n joined connected components $R_1, \ldots, R_n$, obtained after n watershed segmentations, reaches a local minimum in the function $q_i = \frac{\left|R_{(i+\alpha)} - R_{(i-\alpha)}\right|}{|R_i|}$ where α is a user-defined parameter and the operator $|\cdot|$ represents the cardinality of the blob measured in pixels. The MSER detector is by definition covariant to affine transformations.
- *STAR Detector*: This point extractor is also known as Center Surround Extrema (Censure; Agrawal et al. 2008). This approach approximates the Laplacian using bilevel center-surround filters of different shapes such as boxes, octagons, or hexagons. The computation of these filters in combination with integral images allows the detection of interest points in scale-space much faster than SIFT. In our evaluations we used a bilevel star-shaped filter as proposed and implemented in Bradski (2000).

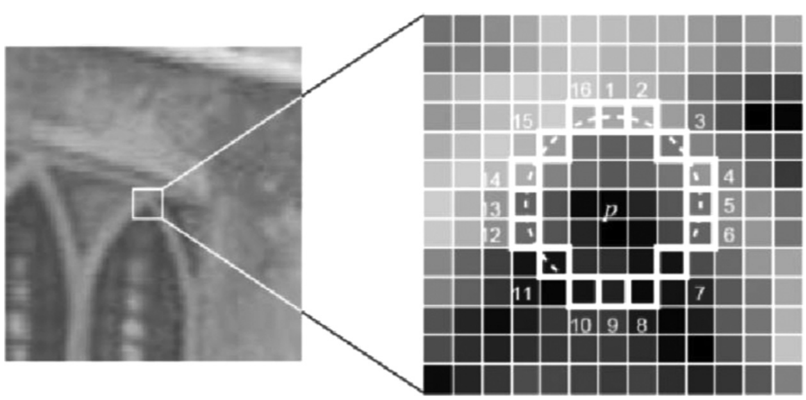


FIGURE 3 FAST local detector (Rosten et al. 2010).

Downloaded by [Iñigo Barandiaran] at 03:41 06 March 2013

- *ORB Detector*: This algorithm proposed in Rublee et al. (2011) is a modified version of the FAST detector for computing orientation during detection step and an efficient computation of a BRIEF-based approach for generating descriptors. This approach tries to merge the rotation and scale invariance of SIFT and the computational efficiency of FAST detector.
- *KAZE Detector*: Introduced in Fernández et al. (2012), this detector proposes a novel multiscale interest point detection, where common linear scale decomposition with Gaussian filtering, used in several approaches such as SIFT, SURF, BRISK, or pyramidal HARRIS, is changed by a nonlinear diffusion filtering. This type of filtering smoothes images similarly to Gaussians but better preserves region boundaries.

EVALUATION

This section shows the results obtained in different tests we carried out following the experimental framework proposed in Barandiaran et al. (2013). This framework allows estimating several performance measures of interest point detectors such as repeatability score, detection accuracy, and computation time. We evaluated the behavior of interest point detectors described in the previous section as implemented using OpenCV Library version 2.4 (Bradski 2000), running entirely in the central processing unit (not using the computer's graphics processing unit). We set all of the specific detectors' parameters to their default values, as suggested by their authors.

Detection Density Evaluation

A detection density test compares the number of interest points that each detector is able to extract. Depending on the specificities of each algorithm, the number of extracted points may vary significantly, even if they are applied on the same image. Furthermore, depending on the image spatial frequencies, the number of detected points can differ. We used three different sets of images having different contents and therefore different textures and spatial frequencies. For example, images from the graffiti data set exhibit well-defined smooth and homogeneous regions, whereas images from the brick data set show highly frequent repeatable patterns. All tests were carried out limiting the maximum number of detections to 6,000.

Table 1 contains the density detection results of all tested detectors over the graffiti, boat, and brick data sets. The ORB and FAST detectors detected the most dense clouds of interest points, followed by KAZE. The ORB detector seemed to always reach the maximum number of detections allowed, in this case 6000, independent of the image content. This tended to generate very close detection of points or clusters, which may have a negative impact in some applications such as camera tracking. Similarly, a high number of

Downloaded by [Iñigo Barandiaran] at 03:41 06 March 2013

TABLE 1 Density Results

	SIFT	SURF	BRISK	ORB	FAST	HARRIS	MSER	STAR	KAZE
Graffiti	1,108	2,505	1,080	6000	5,759	699	555	874	5,209
Boat	1,451	4,088	3,691	6000	4,850	3,426	192	1,923	5,209
Brick	1,821	5,371	1,511	6000	5,458	5,571	1,447	1,461	5,209

detections were obtained with the KAZE detector; however, points detected by KAZE were more uniformly distributed over the image domain than points detected by ORB or FAST. It is worth mentioning that the MSER approach generated the lowest number of detections. A very small number of detections can limit the usefulness of the detectors in some applications such as simultaneous location and mapping (SLAM) or 3D reconstruction, where dense detections are preferable. Finally, we remind the reader that the number of points detected is not the only measure for a successful detector; how discriminative and repeatable they are against some transformations such as geometric or photometric ones is also important.

Invariance to Geometric Transformations Evaluation

In this section we describe the results evaluating the robustness against rotation and scale similarity transformations, affinity transformation, and perspective transformations.

Rotation similarity transformation

In this test we evaluated how different approaches are robust against image rotation. We used the first image from the graffiti data set along with the tool described in Barandiaran et al. (2013) to generate rotated images by applying different angles of in-plane rotation similarity, starting from 0 degrees (same image) to 360 degrees in steps of 7.2 degrees.

The results depicted in Figure 4 show that some detectors such as SIFT or MSER are almost insensitive to in-plane rotation transformation, obtaining an almost constant value of 70% repeatability along the whole transformation range. ORB is also insensitive to transformation but its repeatability values are lower than those of SIFT and MSER, around 55%. Some detectors such as KAZE and, in particular, SURF show high sensitivity to specific in-plane rotation values like 45, 90, 135, 180, 225, or 270 degrees. In the case of SURF we attribute this sensitivity to discretization effects induced by the use of box filters as approximations of the LoG operators. It is worth noticing that the FAST detector, despite its simplicity, obtains good results along the transformation range, generating the best results, together with STAR and KAZE detectors, when image rotation is exactly 180 degrees (upside-down image). The remaining detectors estimate a dominant orientation in the supporting

Downloaded by [Iñigo Barandiaran] at 03:41 06 March 2013

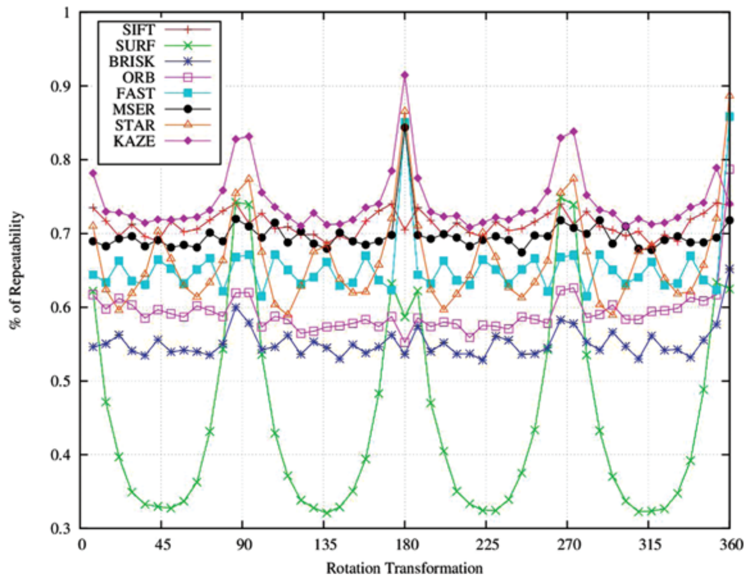


FIGURE 4 Invariance to rotation transformation measured by repeatability results on the first image from the graffiti data set (color figure available online).

region around each interest point, allowing the image to be rotated back or rectified, in order to obtain robustness to rotation transformation. The FAST detector only evaluates some pixels (from 9 to 16) around an interest point without the need for dominant rotation estimation and correction and thus is computationally optimal.

Scale similarity transformation

We again used the first image from the graffiti data set to generate new isotropically scaled views of that image. More precisely, we generated 50 images with a range of scale factors from 0.04 to 2.4. Scale values below 1 indicate augmentation of image structures, and values above 1 indicate a reduction.

Figure 5 plots the repeatability results in this experiment. Clearly, the SIFT detector shows superior results when the value of scale factor transformation is extreme, because it is robust even with scale factors higher than 2. In addition, MSER and BRISK obtained good results, performing better than SIFT for scale factors lower than 1; that is, when images are augmented versions of the original one or the camera is moving closer to the scene, so the apparent size of the objects appears to increase. Finally, it is worth mentioning that the FAST detector is not invariant to scale transformations, given results for repeatability close to 0 when scale factors are outside the range 0.65–1.25.

Downloaded by [Iñigo Barandiaran] at 03:41 06 March 2013

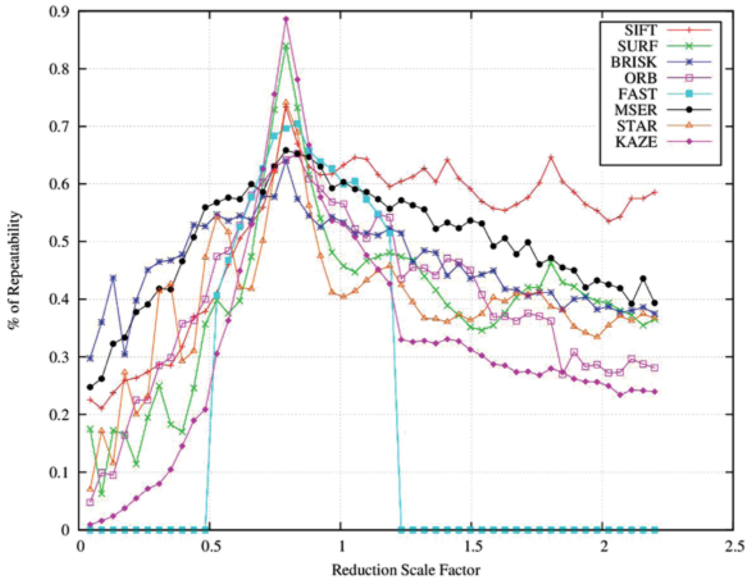


FIGURE 5 Invariance to scale transformation measured by repeatability results on the first image from the graffiti data set (color figure available online).

Downloaded by [Iñigo Barandiaran] at 03:41 06 March 2013

Affine transformation

In addition to the most general transformations (projectivities), affine transformations are the most interesting transformations modeling a change in the camera's viewpoint. They are very useful in several contexts such as SLAM or camera tracking. In this test, we used 50 generated images by applying affine deformation (nonuniform scaling and skew) in the x direction from image 0 to image 25 and then in the y direction from image 26 (most distorted image) to 50 (original image). The results shown in Figure 6 demonstrate that none of the detectors is fully invariant to affine transformation but all of them perform robustly. The KAZE detector obtained the best results, achieving 85% on average along the transformation range.

Perspective transformation

In the following test we evaluated robustness against homography projective transformation. The projective transformation between any two images of the same planar structure in space can be described by a homography transformation. Homography transformation estimation is widely used by the computer vision community in many applications such as image rectification, image registration, camera calibration, or camera pose estimation. In the current test, we used the first four images from the graffiti data set to measure the repeatability score between image 1 (reference) and the other three images. We left

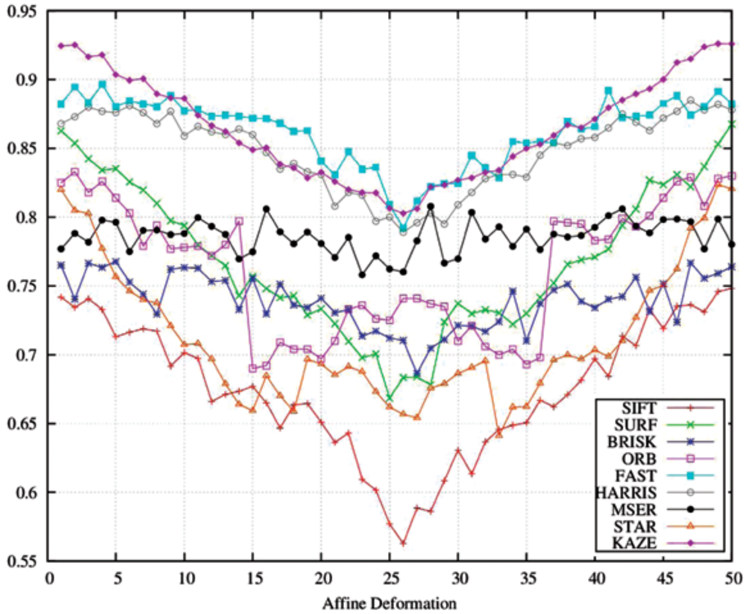


FIGURE 6 Invariance to affine transformations measured by repeatability results (color figure available online).

out the last two images of the data set because the perspective distortion between the reference image and these images was too severe. This distortion limited the applicability of each detector because they were unable to extract a significantly high number of stable interest points; thus the repeatability scores were not very reliable.

The results in Figure 7 show that none of the tested detectors was truly invariant to perspective transformation. BRISK and ORB achieved the best results, followed by KAZE. In general, the repeatability scores in this test were lower than in the rest of tests, meaning that the detectors were very sensitive to perspective transformations and distortions. All of the current approaches propose to extract interest points and descriptors to be invariant to affine geometric transformation because projectivities are too general and thus have too many degrees of freedom. Moreover, when distortion generated by perspective transformation is not very high, this transformation can be locally approximated by an affinity. Therefore, affine invariant detectors such as MSER can be robust against small perspective transformations.

Invariance to Photometric Transformations Evaluation

In addition to geometric transformations, we carried out an evaluation of the robustness of the described interest point detectors against photometric transformations.

Downloaded by [Iñigo Barandiaran] at 03:41 06 March 2013

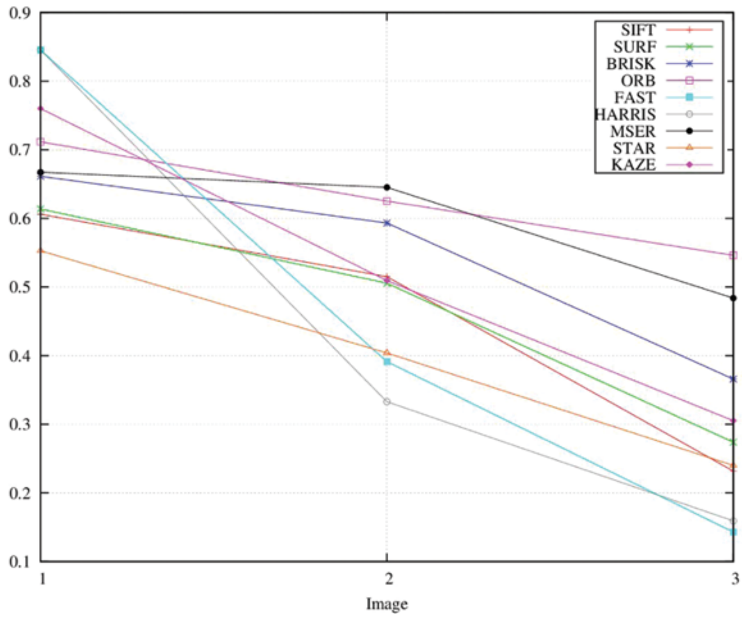


FIGURE 7 Invariance to projectivity transformation results measured by repeatability results (color figure available online).

EXPOSURE PHOTOMETRIC TRANSFORMATION

This test evaluates the robustness of the detectors against variations of light intensity. We used the data set proposed in Barandiaran et al. (2013) consisting of 15 images captured in controlled light conditions. The light was modified from a correct scene exposition to around 4.5 f-stops less of exposure, in steps of 1/3 f-stop. Sample images are shown in Figure 8.

The results obtained with this data set suggest that light intensity variations affect each detector. As the light decreased, the repeatability scores for each detector also decreased. The most stable results were obtained by BRIEF and SURF, followed by MSER. As shown in Figure 9, as the light intensity decreased the number of detections achieved by each detector also decreased with the exception of BRIEF. When the light intensity was reduced

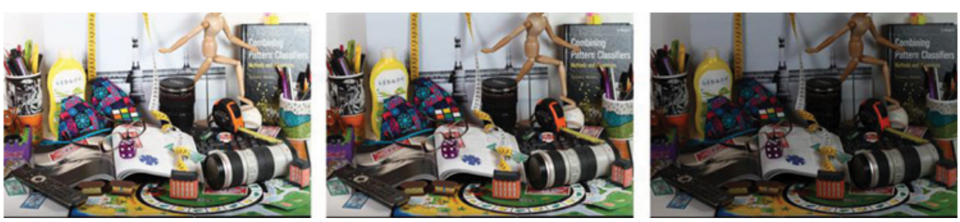


FIGURE 8 Sample images of photometric exposure transformation data set (color figure available online).

Downloaded by [Iñigo Barandiaran] at 03:41 06 March 2013

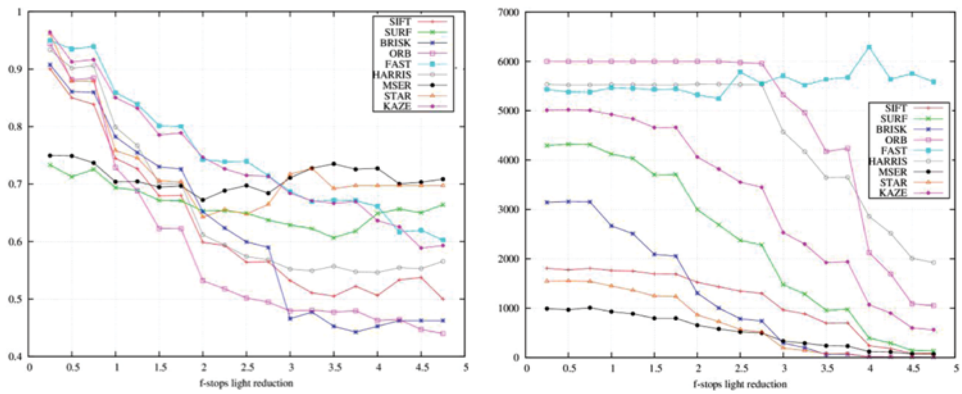


FIGURE 9 Exposure data set repeatability score (left) and number of detections (right) (color figure available online).

Downloaded by [Iñigo Barandiaran] at 03:41 06 March 2013

to around 3 f-stops, the number of detections achieved by each detector was reduced to less than 50% of the total number of detections with correct exposure. As described previously, the BRIEF detector is based on the computation of relative pixel intensity differences. Clearly, this approach is robust and invariant to linear intensity light variations.

Noise photometric transformation

We also evaluated the robustness of interest point detector algorithms against image noise. The current evaluation deals with approaches using image intensity only and thus no color information. We used a data set composed of 15 images that progressively contaminates input image with luminance additive Gaussian distributed noise as shown in Figure 10.

Contrary to the previous experiment, the number of detections for each approach increased as image noise increased. This was due to the addition of spurious data that generated new responses while computing image derivatives. These spurious data caused new false responses (interest points)

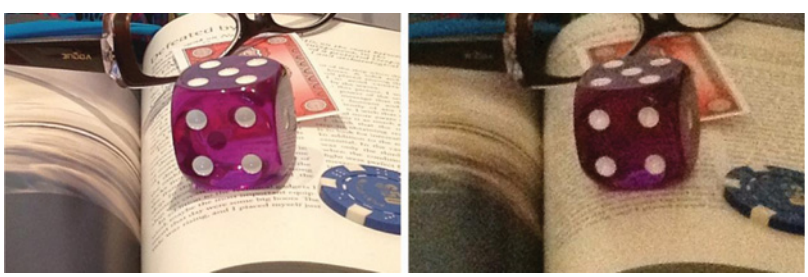


FIGURE 10 Detail of two images with different signal-to-noise ratios (color figure available online).

during the search of local maxima or minima over different scales. Despite the number of detections, clearly these false interest points were not stable; thus, repeatability scores continuously decreased as image noise increased, as depicted in Figure 11. The most stable detector against image noise was BRISK followed by ORB, but all followed the same trend. None of the tested detectors was fully robust to luminance image noise.

Blurring photometric transformation

This test measures the robustness against image blurring. This photometric transformation may occur due to fast camera movements or by a change on the lens focus point. We used a data set consisting of 15 real images where the lens focus point was modified from a perfectly in-focus image to a completely out-of-focus image.

The results for repeatability and detection density are depicted in Figure 12, which shows that as image blurring increased the number of detection decreased; in some cases—for example, using BRISK—this reduction was severe. Each detector uses some type of image blurring, usually through Gaussian functions, prior to interest point detection in either a single or multi-scale approach. The most stable detectors were BRISK, ORB, and SURF. It is worth noting that some approaches such as FAST, SIFT, and HARRIS are very sensitive to this type of transformation and thus showed the worst results in this evaluation.

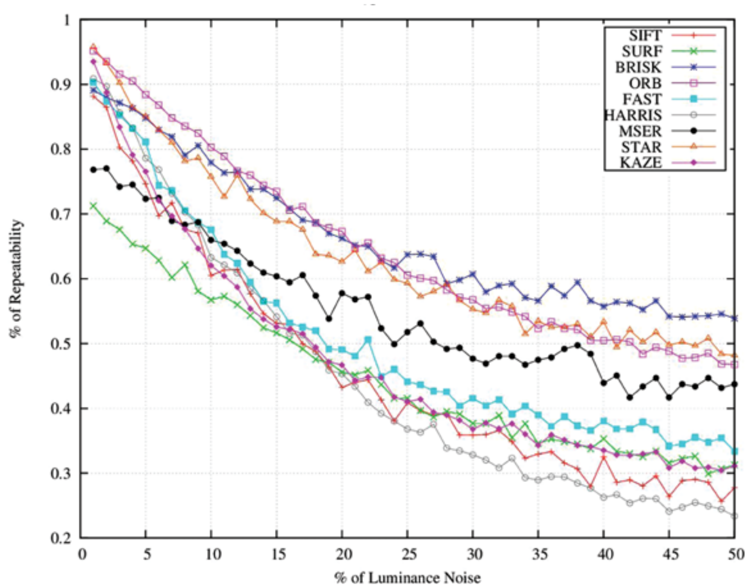


FIGURE 11 Robustness to additive noise measured by repeatability results (color figure available online).

Downloaded by [Iñigo Barandiaran] at 03:41 06 March 2013

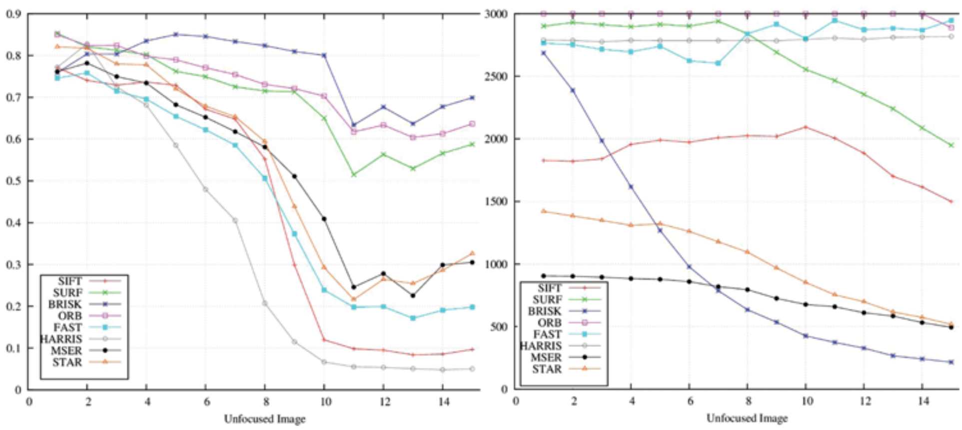


FIGURE 12 Blurring robustness measured by repeatability results (left) and number of detections (right) (color figure available online).

Efficiency Evaluation

This test evaluates the execution times that each detector needs to perform interest point detection in one image. We measured each algorithm by using the six images from the graffiti data set. We used an Intel i5 Quad Core 2.5 GHz with 4 GB of memory.

As in the density test, we set the maximum number of detections to each particular approach to 6,000; thus, each detector was allowed to extract the maximum number of points possible. The results displayed in Figure 13 show clearly that the KAZE feature detector was the slowest method used in this evaluation. The process of computing nonlinear diffusion filtering in several scales is a time-consuming task. The fastest approach for this comparison

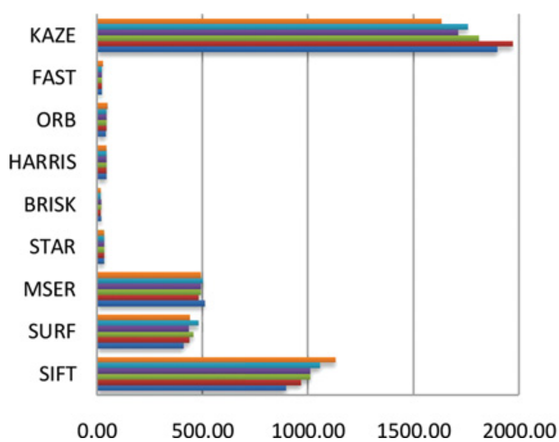


FIGURE 13 Average execution times of interest point detectors (measured in milliseconds; lower is better) (color figure available online).

Downloaded by [Iñigo Barandiaran] at 03:41 06 March 2013

was BRISK, followed very closely by BRIEF and STAR. These approaches along with ORB and HARRIS are suitable for real-time operations.

DISCUSSION

The results in the previous section confirmed that no single interest point detector clearly outperforms the rest of the approaches in all situations. In some tests one particular detector performed better than the others but did not perform as well in other tests. In general, the best approach is the one that provides the best fit for the specific application requirements. For example, the SURF approach performs similar to SIFT, generating more dense interest points, and is computationally faster but suffers from rotation sensitivity, showing irregular results along the rotation transformation range. If our particular application does not expect severe camera or object rotation, SURF can be a perfect alternative to SIFT. Otherwise, if rotations are expected, ORB is a much better option. The ORB detector showed a good trade-off between repeatability in several tests and computational efficiency; however, we observed that the spatial locations of the interest points were usually clustered in very close spatial locations. This spatial clustering may result in the descriptors extracted from such regions not being distinctive enough to effectively perform discriminative matches across images. Conversely, the BRISK detector showed similar robustness measure responses compared to ORB and is computationally faster and, more important, generates much more uniform spatially distributed and stable interest points. The weakest aspect of BRISK in our results is its sensitivity to light intensity changes. Both the number of detections and the repeatability scores decreased drastically as light intensity decreased. Fortunately, the number of computer vision scenarios with such a difference in light exposure is limited, mainly appearing in applications related to outdoor tracking or SLAM, where light conditions are not controlled and may vary significantly from image to image.

Affine transformation robustness is a very important measure, because projective transformations can be locally approximated by an affine transformation. The KAZE and MSER detectors achieved very good robustness results. Despite MSER's robustness to affine transformation, we observed that this approach tended to generate a low number of detections because it requires extensive, well-defined homogeneous regions. This feature can be a serious limitation in many real practical applications.

In addition to robustness to affine transformation, invariance to scale geometric transformation is a critical aspect regarding many interest point matching scenarios, such as camera tracking or object recognition. In this aspect, SIFT is still the best performing algorithm, generating the most stable interest points along different scale factors. Another good performer regarding scale transformation is BRISK, which is much faster than SIFT and thus more suitable for real-time operation.

Downloaded by [Iñigo Barandiaran] at 03:41 06 March 2013

Finally, when real-time operation is a critical requirement, efficient approaches such as FAST, ORB, BRISK, and STAR are the most appropriate. FAST is a very efficient approach, with regard to central processing unit and memory consumption, but is very unstable with regard to scale transformation.

CONCLUSIONS AND FUTURE WORK

In this article we presented an evaluation of different interest point extractors. We reported systematic and exhaustive measurement of their invariance and robustness to several geometric image transformations like similarities (rotations and scaling), affinities, and projectivities. In addition, we evaluated their robustness to photometric transformations such as variations in light exposure or image additive noise. We also evaluated their capability to generate low-level information by measuring the number of points they generate. Finally, we measured their efficiency with regard to computational time requirements.

The choice of the feature detector strongly depends on the application requirements. Overall, we can conclude that recent BRISK detector achieved the best ratio between robustness and efficiency. ORB showed the best performance over rotation transformation, whereas BRISK showed great performance in scale, affine, and projective transformation and was the fastest approach followed by FAST. ORB is a modification of FAST, which does not have an orientation component and does not produce multiscale features. Therefore, FAST was not as accurate as ORB in dealing with rotation and scaling transformations.

Nowadays, efficiency is important, because more and more applications are being migrated to mobile devices, such as the iPad or iPhone. Therefore, approaches similar to FAST or BRISK, which require low computation and memory resources, are useful and promising. The next step is to evaluate some of these algorithms on mobile devices, taking into account that some implementations must be rewritten and optimized to run on specific processor architecture using specific instructions and with several restrictions regarding parallel execution or memory management.

We are now evaluating current state-of-the-art image feature descriptors, which, along with interest point detectors, form the basis of several computer vision applications. Recent feature descriptor approaches like BRISK or FREAK open new possibilities for computer vision applications, such as robust real-time SLAM.

REFERENCES

Agrawal, M., K. Konolige, and M. Blas. "Censure: Center Surround Extremas for Real Time Feature Detection and Matching." In *Proceedings of European Conference on Computer Vision*, 102–115, Marseille, France, October 12–18, 2008.

Downloaded by [Iñigo Barandiaran] at 03:41 06 March 2013

Barandiaran, I., C. Cortes, M. Nieto, and M. Graña. "A New Evaluation Framework and Image Dataset for Key Point Extraction and Feature Descriptor Matching." To Appear in *Proceedings of International Conference on Computer Vision Theory and Application*, Barcelona, Spain, February 21–24, 2013.

Barandiaran, I., C. Paloc, and M. Graña. "Real-Time Optical Markerless Tracking for Augmented Reality Applications." *Journal of Real-Time Image Processing* 5, no. 2 (2010): 129–38.

Bay, H., T. Tuytelaars, and L. Van Gool. "Surf: Speeded Up Robust Features." In *Proceedings of European Conference on Computer Vision*, 404–417. Graz, Austria, May 7–13, 2006.

Bradski, G. "The OpenCV Library." *Dr. Dobb's Journal of Software Tools* 25, no. 11 (2000): 120–126.

Fernández, P., A. Bartoli, and A. J. Davison. "KAZE Features." In *Proceedings of European Conference on Computer Vision*, 214–227. Firenze, Italy, October 7–13, 2012.

Harris, C. and M. Stephens. "A Combined Corner and Edge Detector." In *Proceedings of Alvey Vision Conference*, 14, 147–151. Manchester, England, August 31–September 2, 1988.

Hartley, R. I. and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge, UK: Cambridge University Press, 2004.

Hughes, G. and M. Chraibi. "Calculating Ellipse Overlap Areas." 2011. Available from http://arxiv.org/abs/1106.3787.

Leutenegger, S., M. Chli, and R. Siegwart. "Brisk: Binary Robust Invariant Scalable Keypoints." In *Proceedings of International Conference on Computer Vision*, 2548–2555. Barcelona, Spain, November 6–13, 2011.

Lindeberg, T. *Scale-Space Theory in Computer Vision*. Springer, 1993.

Lowe, D. G. "Object Recognition from Local Scale-Invariant Features." In *Proceedings of International Conference on Computer Vision*, 2, 1150–1158. Corfu, Greece, September 20–25, 1999.

Mair, E., G. D. Hager, D. Burschka, M. Suppa, and G. Hirzinger. "Adaptive and Generic Corner Detection Based on the Accelerated Segment Test." In *Proceedings of European Conference on Computer Vision*, 183–196. Crete, Greece, September 5–11, 2010.

Matas, J., O. Chum, M. Urban, and T. Pajdla. "Robust Wide Baseline Stereo from Maximally Stable Extremal Regions." *Image and Vision Computing* 22, no. 10 (2004): 761–767.

Mikolajczyk, K. and C. Schmid. "Indexing Based on Scale Invariant Interest Points Computer Vision, 2001." In *Proceedings of the Eighth Annual IEEE International Conference on Computer Vision*, Vol. 1, 525–31. IEEE, Vancover, Canada, July 7–14, 2001.

Mikolajczyk, K. and Schmid, C. "An Affine Invariant Interest Point Detector." In *Proceedings of European Conference on Computer Vision*, 128–142. Copenhagen, Sweden, May 27–June 4, 2002.

Mikolajczyk, K., T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaalitzky, T. Kadir, and L. V. Gool. "A Comparison of Affine Region Detectors." *International Journal of Computer Vision* 65, no. 1 (2005): 43–72.

Downloaded by [Iñigo Barandiaran] at 03:41 06 March 2013

Moravec, H. P. “Towards Automatic Visual Obstacle Avoidance.” In *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, 2, 584. Massachusetts, USA, August, 1977.

Rosten, E., R. Porter, and T. Drummond. “Faster and Better: A Machine Learning Approach to Corner Detection.” *Pattern Analysis and Machine Intelligence* 32, no. 1 (2010): 105–19.

Rublee, E., V. Rabaud, K. Konolige, and G. Bradski. “ORB: An Efficient Alternative to SIFT or SURF.” In *Proceedings of International Conference on Computer Vision*, 2564–2571. Barcelona, Spain, November 6–13, 2011.

Tuytelaars, T. and K. Mikolajczyk. “Local Invariant Feature Detectors: A Survey.” *Foundations and Trends® in Computer Graphics and Vision* 3, no. 3 (2008): 177–280.

Viola, P. and M. J. Jones. “Robust Real-Time Face Detection.” *International Journal of Computer Vision* 57, no. 2 (2004): 137–54.

Downloaded by [Iñigo Barandiaran] at 03:41 06 March 2013