Repeatability Evaluation of Keypoint Detection Techniques in Tracking Underwater Video Frames

Ghulam-Sakhi Shokouh, Baptiste Magnier, Binbin Xu, Philippe Montesinos

Repeatability Evaluation of Keypoint Detection Techniques in Tracking Underwater Video Frames

Ghulam Sakhi Shokouh[0000−0003−2561−7317], Baptiste Magnier[0000−0003−3458−0552], Binbin Xu[0000−0002−0822−5250], and Philippe Montesinos[0000−0003−3741−8702]

EuroMov Digital Health in Motion, Univ Montpellier, IMT Mines Ales, Ales, France
{ghulam-sakhi.shokouh, baptiste.magnier, binbin.xu, philippe.montesinos }@mines-ales.fr

Abstract. Keypoint(s) or corners, as a stable feature possessing the defined characteristics of a robust point of interest remain an active research field for machine vision researchers due to its applications in motion capture, image matching, tracking, image registration, 3D reconstruction, and object recognition. There exist different techniques for keypoint detection; this paper focuses on direct computation on the gray-level analysis of interest point detection because of its straightforward implementation. In this contribution, an objective comparison of 12 state-of-the-art keypoint detection techniques; an application to feature matching have been executed in the context of underwater video sequences. These videos contain noise and all geometric and/or photometric transformations. Experiments are led on 5 videos containing in total 10 000 frames, evaluating the repeatability of the keypoints detectors. These detectors are evaluated on these complex videos by computing statistics-based repeatability, but also as a function of the Zero-Mean Normalized Cross-Correlation (ZNCC) scores.

Keywords: Corner detection · filtering · repeatability · underwater imagery · keypoint detection · AQUALOQ dataset · ZNCC.

1 Introduction and Motivation

The importance and interest in keypoint detection (i.e, corner or junction as a stable interest point) in a digital image lies notably in its application in image matching, tracking, motion estimation, panoramic stitching, object recognition, and 3D reconstruction. Image matching through feature tracking is extensively used in many real-time applications including autonomous driving, security surveillance, and manufacturing automation [18]. Corner detection techniques can be effectively applied in these applications depending on their repeatability ratio. The reason for the corner detection's wide range of applications is that the corner is easier to localize than other low-level features such as edges or lines, particularly taking into consideration the correspondence problems (e.g., aperture problem in matching). Hence, an objective evaluation of the frequently

applied corner detection techniques by direct computation on the gray-level analysis relating to their real-time application is primarily invaluable, an example is available in [7].

The image matching and feature tracking in complex real-time scenes such as underwater videos are extremely challenging [5]. In this type of image sequence, concerning all types of image transformation (rotation, scale, affine transformations, translation, etc.), photometric transformation (illumination, occlusion, clutter, etc.), and various types of noises plus moving particles in different directions, the robustness of interest point, can be truly evaluated both objectively and visually. Repeatability is the main evaluation metric widely used for interest point matching, where the obtained points must be independent of varying image conditions [18][15][12][14]. In this work, the repeatability rate of the 12 commonly applied corner detection operators are objectively evaluated on the challenging underwater video dataset [5].

In the literature, several approaches for detecting corners and junctions in digital images have been developed: (i) involving contour chains [2], (ii) using templates [17][23] or, (iii) by image filtering techniques. Mainly, the corner detection operators via the direct computation on the gray-level analysis corresponding to the label (iii) can be categorized in three general approaches: Hessian based [3], curvature analysis [10][26][4][22][24][1], and structure tensor based [6][13][8][19][16][9]. These methods are easily developed by image filtering because they involve only image convolutions horizontally and vertically. Therefore, they can be implemented with less computational time on different devices, see details in [7]. This paper is devoted to an extensive evaluation of filtering-based corner detection methods via repeatability performance measurement in video sequences consisting of frames with different types of transformations. It is to mention that the terms corner, junction, salient point, keypoint, and interest point are used synonymously in this work.

2 Studied Keypoint Detectors by Gray-level Direct Computation

In this section, a set of corner detection techniques including the general scheme and the related parameters have been investigated. There are different approaches to determining the cornerness measure by direct computation using filtering techniques. Generally, in image filtering, the first or second derivatives may be utilized to determine corners in an image. Considering a gray-level image I, its partial derivatives are:

- $I_x = \frac{\partial I}{\partial x}$, the 1st image derivative along the x axis,
- $I_y = \frac{\partial I}{\partial y}$, the 1st image derivative along the y axis,
- $I_{xx} = \frac{\partial^2 I}{\partial x^2}$, the 2nd image derivative along the x axis,
- $I_{yy} = \frac{\partial^2 I}{\partial y^2}$, the 2nd image derivative along the y axis,
- $I_{xy} = \frac{\partial^2 I}{\partial x \partial y}$, the crossing derivative of I.

These image derivatives can be calculated by convolving the image with the [1 0 -1] or the [1 0 -2 0 1] masks in the x and/or y directions for the first and second derivatives, respectively [20]. The first derivatives are useful for the detection of step and ramp edges, whereas the second derivatives are convenient for the contour extraction of types: line, roof edges as ridge/valley features. Regarding the image surface, corners are defined as the curvature extremum along the edge line [16]. Usually, approaches to detect directly corners on the gray scale level use filtering techniques by combining image derivatives of the 1st and 2nd order, then by computing the Hessian matrix, the curvature or the structure tensor. All the necessary technical details of the studied corner detection methods including mainly the formula, denomination, parameter(s), and name of the authors are listed by year of publication in Tab. 1.

2.1 Determinant of the Hessian Matrix

Mathematically, the Hessian matrix ($\mathcal{H}$) indicates significant values near edges, through which corners can be estimated by the large variations of intensity values in both x and y directions. Indeed, $\mathcal{H}$ represents a square matrix of 2nd order partial derivative of image intensity; it is often computed and is useful in feature detection and characterization:

$$\mathcal{H} = \begin{pmatrix} I_{xx} & I_{xy} \\ I_{xy} & I_{yy} \end{pmatrix}. \tag{1}$$

In that respect, Beaudet [3] uses the image 2nd derivative for calculating the determinant of $\mathcal{H}$, which is related to Gaussian curvature of the image surface [11]. The computation is straightforward because it combines only three 2nd image derivatives. Even though this technique is rotation invariant, it is sensitive to noise and unstable against scale changes.

2.2 Curvature Analysis

Technically, these techniques are based on the change of gradient direction along an edge contour and/or image surface curvature. They can be easily computed by the combination of the image derivatives of 1st and 2nd order. The pioneer work in this category was originally led by Kitchen and Rosenfeld (KR) who defined the cornerness measure for each pixel intensity based on the change of 2nd order gradient direction along the edge weighted by the local gradient magnitude [10]. Theoretically, the gradient feature vector ∇I is normal to the edge and hence projecting the change of gradient direction along the edge and multiplying the result by the local gradient magnitude $|\nabla I|$ results the final cornerness measure. Inspired by this initial contribution of Kitchen and Rosenfeld [10], other related techniques were developed by Zuniga and Haralick [26], Blom *et al.* [4], Wand and Brady [22]. Zheng *et al.* [24] and Achard *et al.* [1] uses a smoothed image with a Gaussian of parameter ρ and then the combination of its derivatives with the derivatives of the non-smoothed image. The cornerness measure for each technique is listed in the Tab. 1.

Table 1. Cornerness measure formulas computed from image derivatives. Here, $\overline{I_x}$, $\overline{I_y}$, $\overline{I_{xx}}$, $\overline{I_{yy}}$ and $\overline{I_{xy}}$ denote the convolutions with a Gaussian with a standard deviation $\sigma > 0$ of images derivatives I_x, I_y, I_{xx}, I_{yy} and I_{xy} respectively. As a reminder, (λ_1, λ_2) represents the eigenvalues of the structure tensor $\mathcal{M}$.

Name	Cornerness Measure/Formula	Parameter(s)	Reference(s)	Year
DET	$DET(\mathcal{H}) = I_{xx}I_{yy} - I_{xy}^2$	-	[3]	1978
KR	$\frac{I_x^2 I_{yy} - 2 \cdot I_x I_y I_{xy} + I_y^2 I_{xx}}{I_x^2 + I_y^2}$	-	[10]	1982
ZH	$\frac{KR(I)}{\lvert\nabla I\rvert} = \frac{I_x^2 I_{yy} - 2 \cdot I_x I_y I_{xy} + I_y^2 I_{xx}}{\left(I_x^2 + I_y^2\right)^{3/2}}$	-	[26]	1983
F	$\frac{Det(\mathcal{M})}{Trace(\mathcal{M})} = \frac{\lambda_1\lambda_2}{\lambda_1 + \lambda_2} = \frac{\lambda_1\lambda_2}{\lvert\nabla\overline{I}\rvert^2} = \frac{\overline{I_x}^2\overline{I_y}^2 - \overline{I_x I_y}^2}{\overline{I_x}^2 + \overline{I_y}^2}$	ρ	[6], [13]	1987
HS	$Det(\mathcal{M}) - k \cdot (Trace(\mathcal{M}))^2 = \overline{I_x}^2\overline{I_y}^2 - \overline{I_x I_y}^2 - k \cdot \left(\overline{I_x}^2 + \overline{I_y}^2\right)^2$	ρ, k	[8]	1987
BB	$\lvert\nabla I\rvert \cdot KR(I) = I_x^2 I_{yy} - 2 \cdot I_x I_y I_{xy} + I_y^2 I_{xx}$	-	[4]	1992
Ro	$Det(\mathcal{M}) = \lambda_1\lambda_2 = \overline{I_x}^2\overline{I_y}^2 - \overline{I_x I_y}^2$	ρ	[16]	1994
KLT	$Min(\lambda_1, \lambda_2)$	ρ	[19]	1994
RTC	$\frac{(1 + I_x^2) I_{yy} - 2 \cdot I_x I_y I_{xy} + (1 + I_y^2) I_{xx}}{\left(1 + I_x^2 + I_y^2\right)^{3/2}}$	-	[22]	1995
GD	$I_x^2 I_{yy}^2 + I_y^2 I_{xx}^2 - \frac{\overline{I_x}^2\overline{I_{yy}^2} + \overline{I_y}^2\overline{I_{xx}^2}}{\left(\overline{I_x}^2 + \overline{I_y}^2\right)^2} \cdot \left(I_x{}^2 + I_y{}^2\right)^2$	ρ	[24]	1999
ABD	$\frac{I_x^2\overline{I_y}^2 + I_y^2\overline{I_x}^2 - 2 \cdot I_x I_y \overline{I_x I_y}}{\overline{I_x}^2 + \overline{I_y}^2}$	ρ	[1]	2000
KZ	$\frac{1}{\left(\lambda_1^{-p} + \lambda_2^{-p}\right)^{1/p}}$	ρ, $p > 0$	[9]	2005

2.3 Structure Tensor

The third group uses the symmetric structure tensor $\mathcal{M}$:

$$\mathcal{M} = \begin{pmatrix} \overline{I_x^2} & \overline{I_x I_y} \\ \overline{I_x I_y} & \overline{I_y^2} \end{pmatrix}, \tag{2}$$

where $\overline{\bullet}$ indicates convolution with a low-pass filter; here a Gaussian filter of standard deviation ρ is considered (see Eq. (3) with a parameter ρ instead of σ).

The structure tensor is derived from the gradient structure tensor, which is achieved through, computing the Cartesian product of the gradient vector with itself at each point of the image. Spatial averaging of this tensor, usually with a Gaussian filter, leads to the structure tensor. As to note that averaging is needed as the plain gradient tensor has only one non-zero eigenvalue and therefore represents only innately edge features. Spatial averaging, here tied to ρ parameter, distributes this information over a neighborhood, and points that receive contributions/input from edges with different orientations will have two positive eigenvalues, which allows them to be recognized as intrinsically 2D. Eigenvectors of the gradient structure tensor indicate local orientation, whereas

eigenvalues λ_1 and λ_2 give the strength or magnitude as a measure of eigenvalues in flat regions are very small (negligible), in the edges λ_1 or λ_2 is small depending on the horizontal or vertical edge, and noticeably both values λ_1 and λ_2 are large in corner points.

Based on this assumption, various corner measurement formulations have been proposed; they are listed and denominated in the Tab. 1 and summarized here:

- Fostner [6] and Noble [13] use an auto-correlation matrix $\mathcal{M}$ with the function F to identify salient points, which converges toward the point closest to all the tangent lines of the corner in a neighborhood and is a least-square solution. The function combines the eigenvalues, aiming to classify the keypoints from other types of local features.
- Harris and Stephens [8] also named as Plessey operator is based on principal curvature of local auto-correlation using first order derivative. This operator's response is theoretically isotropic, but often computed in anisotropic way. This cornerness measure HS yields two positive values at corners and two negative values in the case of straight edges. HS and F methods differ only in their criterion determination because the Fostner algorithm disregards the parameter k ($k > 0$) introduced in Harris and Stephens by computing a fraction.
- Rohr used a parametric model fitting as a point of maximal curvature in the direction of the steepest slope [16]. He convolved an analytical corner model with a Gaussian and adjusted the parameter of the model by a minimization technique to have the model near the observed signal. It is remarkable that this algorithm corresponds to HS technique for a value $k = 0$.
- Shi and Tomasi estimated that the corners are primitives which remain more stable for tracking, contrary to other features [19]. Consequently, the minimum eigenvalue between λ_1 and λ_2 of matrix $\mathcal{M}$ is conserved for a salient point along a video; then this detector led to the well-known KLT (Kanade-Lucas-Tomasi) feature tracker.
- Kenney *et al.* [9] combines λ_1 and λ_2 with a cornerness measure which is constrained by the numbers of required axioms to satisfy. The axioms mainly formulate the isotropy condition (rotation invariant corner), orthonormality of the matrix M, constant eigenvalues relating to the norm, and finally definition of the maximum value of the isotropic point over the set of constant eigenvalues. As detailed in [9], KZ detector technique is equivalent modulo for the choice of a suitable matrix norm and a normalization constant to:
  - ☐ F when $p = 1$,
  - ☐ KLT when $p \to +\infty$,
  - ☐ $\sqrt[p]{2}$R for $p \to 0$.

  In our experiments, p is fixed to 2.

These six corner detection techniques have in common the tensor $\mathcal{M}$, which is tied to the same low-pass filter parameter, here denoted ρ: the standard deviation of the Gaussian. This ρ value is identical for each compared technique in the experiment presented in the next section.

2.4 Detection of a Corner: Final Step

For corner extraction, the final step concerns the binarization of the detected salient points or corners. The obtained features from the techniques presented in subsections 2.1–2.3 and listed in the Tab. 1 compute cornerness measure by applying non-maximum suppression where the local maxima are tied to the corner positions (here a window of size 15×15 is chosen to avoid too close keypoints). Eventually, corner points are highlighted by thresholding the extracted points or by setting the number of corner points to be detected objectively (this last solution was adopted in our experiments).

3 Keypoint Repeatability Assessments

Repeatability measure is the defacto standard and is commonly applied for the performance characterization of salient point detectors [21][15]. The repeatability rate measures the detector's robustness in being able to detect the same features in the condition of image perturbations (e.g., a corner detector that is robust in the condition of image perturbation is rated as a highly repeatable detector). To pursue a vigorous evaluation of techniques detailed and nominated in Tab. 1, our experiments are carried out on a specific database containing strong perturbations, as reported in Tab. 2 and detailed in the next subsection.

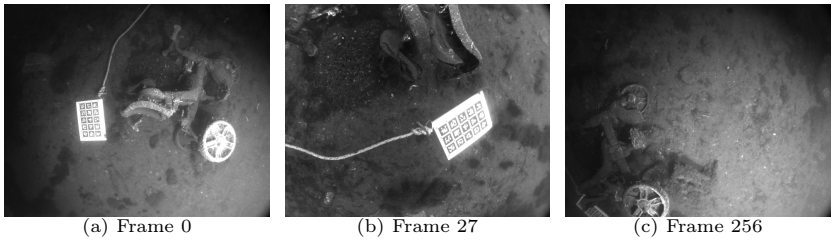
Table 2. Experimental protocols of selected videos in the AQUALOC dataset.

Sequence №	№ of Frames	Description of image transformation
Video sequence 1	1000	rotation, affine, illumination, scale
Video sequence 2	2000	rotation, affine, illumination, perspective
Video sequence 3	2000	rotation, illumination, homogeneous, scale
Video sequence 4	2000	rotation, illumination, occlusion, translation
Video sequence 5	3000	rotation, illumination, affine, scale, clutter

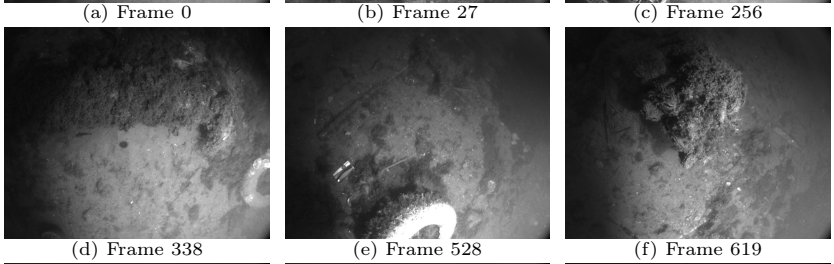
3.1 Experimental Protocol

In this assessment, the repeatability rate is considered as the percentage of the total number of points that are repeated between two subsequent video frames in the AQUALOQ dataset[1]. This dataset is an underwater video sequence dedicated to the localization of underwater vehicles navigating close to the seabed. These videos have been recorded from remotely operated vehicles equipped with a monocular monochromatic camera. The image data is complex in consisting of all types of geometric and photometric transformation plus different types

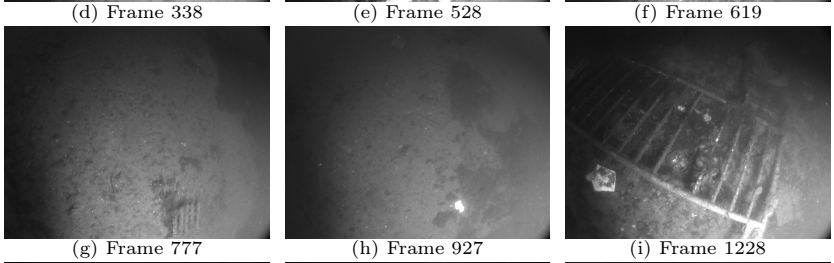
[1] The AQUALOC Dataset [5] is available at: https://www.lirmm.fr/aqualoc/



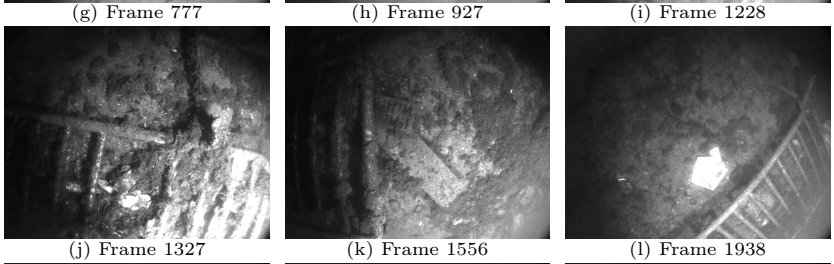
(a) Frame 0 (b) Frame 27 (c) Frame 256



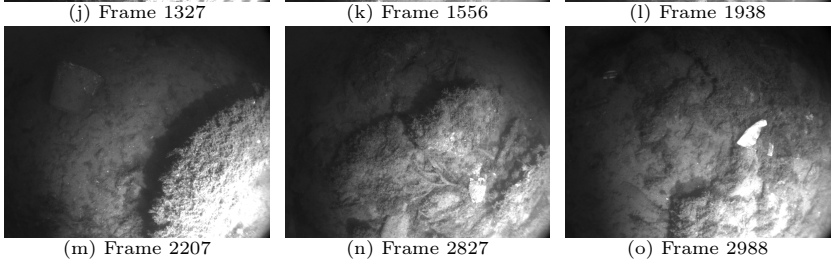
(d) Frame 338 (e) Frame 528 (f) Frame 619



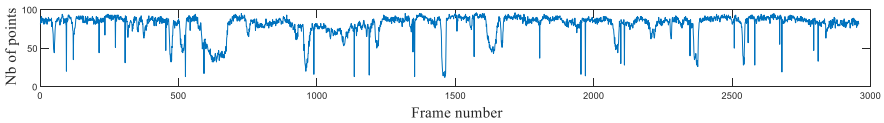
(g) Frame 777 (h) Frame 927 (i) Frame 1228



(j) Frame 1327 (k) Frame 1556 (l) Frame 1938



(m) Frame 2207 (n) Frame 2827 (o) Frame 2988



(p) Number of detected points by Ro as a function of the frame number along the video 5.

Fig. 1. Selected images of video 5 of AQUALOC dataset [5] for visual explanation of all types of transformation and noises, images of size 512×640.

of noises and concurrent moving particles moving in different directions as illustrated in the Fig. 1. Nevertheless, the frame movements are very smooth, denoting a small inter-frame distance. For our statistical experiments, we randomly selected 5 videos with different numbers of frame sequences. With a total of 10 000 frames, video 1 includes 1000 frames, videos 2, 3 and video 4 contain 2000 frames, finally, video 5: 3000 frames, as listed and detailed in the Tab. 2.

In order to remove the noise in the image and obtain more relevant keypoints, the zeroth order two-dimensional Gaussian kernel G is used for regularization by convolution with the image [25]. Its equation is given by:

$$G(\sigma, x, y) = \frac{1}{2\pi\sigma^2} \cdot \mathrm{e}^{-\frac{x^2+y^2}{2\sigma^2}} \text{ with } \sigma \in \mathbb{R}_+^*,\ (x, y) \in \mathbb{R}^2, \tag{3}$$

where σ represents the standard deviation of the Gaussian G and (x, y) the pixel coordinates. Hence, the images are smoothed with G of parameter $\sigma = 1$ before applying keypoint detector techniques. Regarding these 12 keypoint detectors, detailed in Section 2 and listed in Tab. 1, all of them detect the 100 best points per frame. The choice of ρ parameter value for the detectors consisting of ρ parameter, is usually made empirically because a too large value can delocalizes the keypoint position and will "disrupt" the repeatability. Indeed, when the ρ value increases from certain threshold, the keypoints can get misplaced increasingly (see example in [15]) and some of them could be merged. Meanwhile, too small ρ value is a restricting threshold, limiting the detection of structure and will result the low repeatability ratio for matching. Therefore, for each detector, the ρ value is estimated in the Sec. 3.2.

3.2 Evaluation via ZNCC Process

Tracking by matching the features is defined as obtaining the features (i.e., keypoints) in the first frame I_1 of the video sequence and then finding the corresponding pairs of points in the subsequent frame I_2. After detecting a feature point in I_1, a feature point in I_2 is generally estimated and located by computing the intensity variation between a patch in I_1 and patches in I_2. There are several straightforward metrics to estimate the similarity between the two intensity patches, such as Sum of Squared Differences (SSD), Sum of Absolute Difference (SAD), Normalized Cross-Correlation (NCC), and Zero-Mean Normalized Cross-Correlation (ZNCC). For our assessments, the ZNCC is chosen as the optimal evaluation metric in matching and tracking, because it is more precise as being less sensitive to proportional changes of intensity:

$$\mathrm{ZNCC}(\Omega_1, \Omega_2) = \frac{\sum_i^N \sum_j^N [(\Omega(i,j) - \mu_1).(\Omega_2(i,j) - \mu_2)]}{N^2 \cdot \sigma_1 \cdot \sigma_2}, \tag{4}$$

where, Ω_1 and Ω_2 correspond to the frame patches of size $N \times N$ pixels, (μ_1, σ_1) and (μ_2, σ_2) are the mean and standard deviations of the intensities of the patches Ω_1 and Ω_2 respectively. In case where a keypoint is calculated in a homogeneous region: $\sigma_1 = 0$ or $\sigma_2 = 0$, consequently Eq. (4) does not compute

a value in the desired range, so $\text{ZNCC}(\Omega_1, \Omega_2) = 1 - |\mu_1 - \mu_2|$. The ZNCC computes the similarity measures between the two equally sized image patches (I_1, I_2), and gives a scalar in the range $[-1; 1]$. The value/score between $[0; 1]$ indicates the ratio of positive correlativity of the features. The closer to 1 the score is, the similarity between the patches is. As there are small displacements between each frame in the studied dataset, consequently, the ZNCC descriptor is applicable for matching because patches spatially close to another patch in the subsequent frame will obtain a positive ZNCC score.

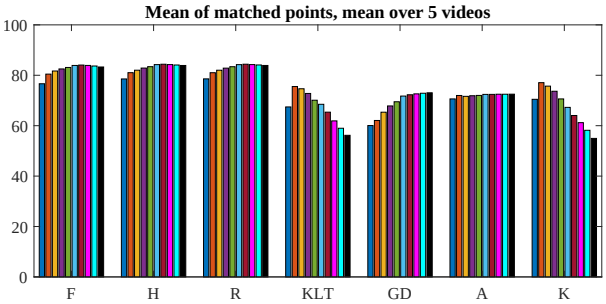
In the evaluation process, for scoring the repeatability ratio of each studied feature detector, three statistical metrics are first computed, namely: (i) mean of matched points (percentage of matched points exactly), (ii) Standard deviation (Std) of matched points and (iii) ZNCC mean for each frame. Thereafter, in order to obtain an evaluation as objective as possible, the feature point detectors having a ρ parameter are compared by varying this parameter. Indeed, ρ is increasing from 0.5 to 4.5 by a step of 0.5 and the repeatability ratio is estimated for each scale, see Sec. 3.2. Consequently, the Table 3 reports the mean and Std of matched points as a function of the video number for each detector (and best ρ). To complete the evaluation, a final score is computed in Eq. (5) to estimate the reliability of the detectors as a function of 3 entities:

- $\mathcal{M}_{Matched}$: the usual mean of matched points among 100 detected points along the 5 videos.
- $\text{S}td_{Matched}$: the usual standard deviation of matched points along the 5 videos.
- $\mathcal{M}_{ZNCC}$: the mean of ZNCC scores for all the matched points along the 5 videos.

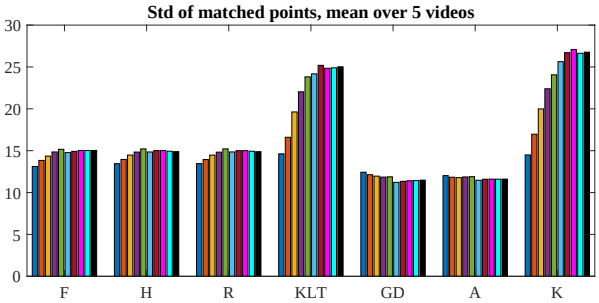
These 3 values are displayed in Fig. 2(a)-(d) as a function of the ρ values regarding the 7 feature detectors: F, H, R, KLT, GD, A and K. Usually, it is worth noting that the $\text{S}td_{Matched}$ and $\mathcal{M}_{ZNCC}$ score can be inversely proportional to the number of detected points; an example, a detector can match few number of points by having a low $\text{S}td_{Matched}$ value, as ZH in Tab. 3. Furthermore, $\mathcal{M}_{Matched}$, $\text{S}td_{Matched}$ and $\mathcal{M}_{ZNCC}$ are normalized and the final score is computed by:

$$\mathcal{T}_{\mathcal{S}} = \frac{a \cdot \mathcal{N}(\mathcal{M}_{Matched}) + b \cdot \mathcal{N}(\text{S}td_{Matched}) + c \cdot \mathcal{N}(\mathcal{M}_{ZNCC})}{a + b + c}, \tag{5}$$

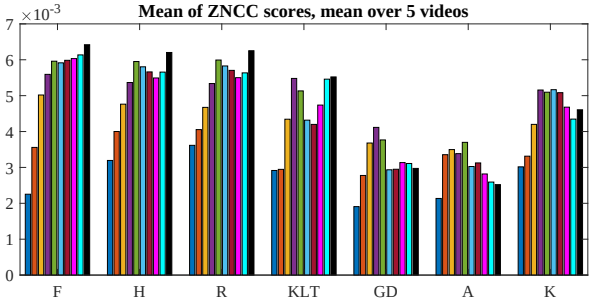
where $\mathcal{N}$ represents the normalization function, and (a, b, c): 3 positive coefficients such that $(a + b + c) = 1$. In our experiments, $a = 0.4$, $b = 0.3$ and $c = 0.3$ to correspond to weights such that the mean of detected points remains the main entity. The Fig. 2(d) reports different scores as a function of the ρ variation. This total score is also computed for the detector without ρ parameters, namely: DET, KR, ZH, BB and RTC. Finally, all the statistics are reported in the Tab. 4. The more the $\mathcal{T}_{\mathcal{S}}$ score is close to 1, the more the feature detector is qualified as suitable for repeatability. On the contrary, a score close to 0 indicates a low reliability of a detector.



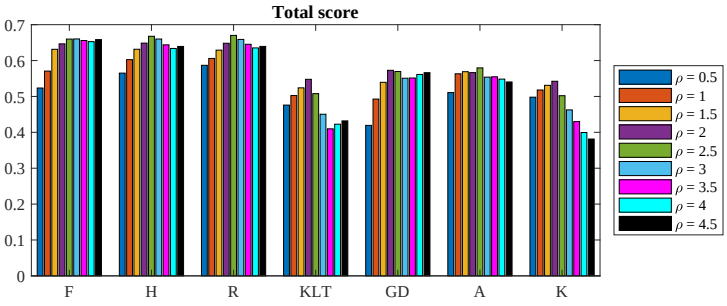
(a) Mean of matched points $\mathcal{M}_{Matched}$ among 100 detected points



(b) Standard deviation of matched points $Std_{Matched}$



(c) Mean of ZNCC scores $\mathcal{M}_{ZNCC}$ for all the matched points $\mathcal{M}_{Matched}$ in (a)



(d) Total score $\mathcal{T}_S$ computed as in Eq. (5)

Fig. 2. Matched and repeatability statistics as function of ρ values for 7 filtering techniques averaged on 5 videos: ρ values in the range of 0 to 4.5

Table 3. Percentage and Std of detected points per video

Detector	Video 1		Video 2		Video 3		Video 4		Video 5	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
DET	81.9	7.6	76.4	14.6	80.1	13.1	68.7	15.7	77.9	11.9
KR	76.9	7.1	70.6	13.9	74.6	12.0	64.6	14.4	71.5	11.0
ZH	8.3	5.1	10.3	9.0	11.7	6.8	10.1	8.5	12.2	9.1
F	**90.2**	8.7	81.8	17.0	77.0	17.9	**88.8**	11.6	**84.5**	15.4
HS	90.1	8.6	**82.0**	17.0	**77.2**	17.7	88.5	12.4	82.9	16.5
BB	84.6	7.9	77.9	15.3	81.5	13.5	10.1	8.6	79.3	12.7
Ro	90.1	8.6	**82.0**	17.0	**77.2**	17.7	88.5	12.4	82.9	16.5
KLT	81.0	9.5	70.1	23.1	57.3	29.1	86.5	12.0	71.6	27.5
RTC	82.2	7.7	76.6	14.8	80.6	13.2	71.3	16.2	78.2	12.0
GD	64.4	11.8	68.8	12.9	64.6	13.2	71.2	9.2	69.6	12.0
A	78.1	7.0	71.1	13.4	67.1	13.8	73.8	10.0	72.0	12.5
KZ	86.3	9.9	71.7	23.7	57.8	29.6	87.4	12.1	72.2	27.8

Scale ρ parameter fitting Depending on the ρ parameter, the detectors F, H, and R shows the highest stability for different scale ratio as presented in the Fig. 2(d) total score $\mathcal{T}_{\mathcal{S}}$. Here, the detectors, KLT and K total score decreases for ρ values greater than 2.5, whereas the detectors GD and A produce average total score values. Correspondingly, the statistics tied to the best scores regarding ρ parameter of these detectors are reported in the Table 4, as for statistics in the Tab. 3. It is important to note that values are rounded in Tab. 3; whereas original/exact values serve for Tab. 4.

Usual statistical evaluation The mean ZNCC values for each operator demonstrate the cumulative average similarity or correlation of only the corresponding patches in the given image sequences; they are also reported in the Tab. 4. The obtained values are all positive values, implying certain similarities and measures of repeatability of the corresponding matched points. Nevertheless, the mean ZNCC values are a function of matched points per detected points. As an example, the mean ZNCC values for the two detectors Ro and ZH (best versus worse detectors) correspond to the number of matched points per total number of detected points. It is to note that, a solid interpretation of the obtained values is complicated as each different operator's performances with different types of image transformation, occlusion and noises vary widely, as illustrated in Fig. 1. To be recalled again, for this experiment the displacement of frames are smooth and small, so keypoints in consecutive frames should not be too far spatially from the successive frames in the experimental videos. As an example, the detector DET shows a high percentage of matched corner points in the first video sequences; however, in the video 4, drastic changes in scores appear relating to the different transformations (81.9 against 68.7 in Tab. 3). Hence, the repeatability ratio is always dependent on the type of image transformation. Another example concerns a group of structure tensor based techniques having ρ scale parameter. Even though the detector Ro, HS, and F obtained the optimal match-

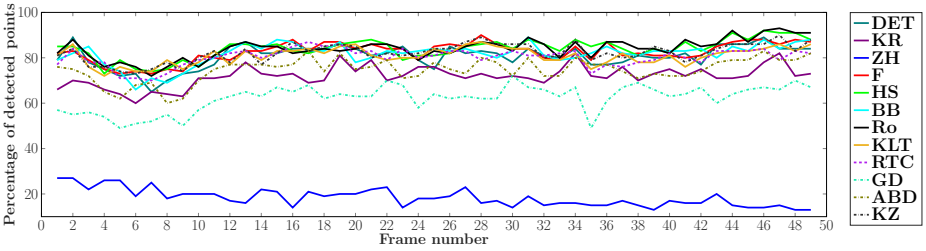
Table 4. Main repeatability statistical scores of matched points, mean ZNCC (rounded here) and final conclusive scores $\mathcal{T}_S$ in Eq. (5).

Detector	DET	KR	ZH	F	HS	BB	Ro	KLT	RTC	GD	A	KZ
ρ parameter	-	-	-	3	2.5	-	2.5	2	-	2	2.5	2
$\mathcal{M}_{Matched}$	76.62	71.11	10.92	83.90	83.41	66.15	83.40	72.79	77.36	67.83	72.01	73.66
$Std_{Matched}$	13.01	12.04	8.09	14.76	15.22	12.06	15.21	22.03	13.19	11.85	11.89	22.39
$\mathcal{M}_{ZNCC}$	0.008	0.007	0.15	0.006	0.006	0.010	0.006	0.005	0.008	0.004	0.003	0.005
$\mathcal{T}_S$, see Eq. (5)	0.34	0.32	0.07	0.66	0.66	0.30	0.67	0.55	0.35	0.57	0.58	0.54

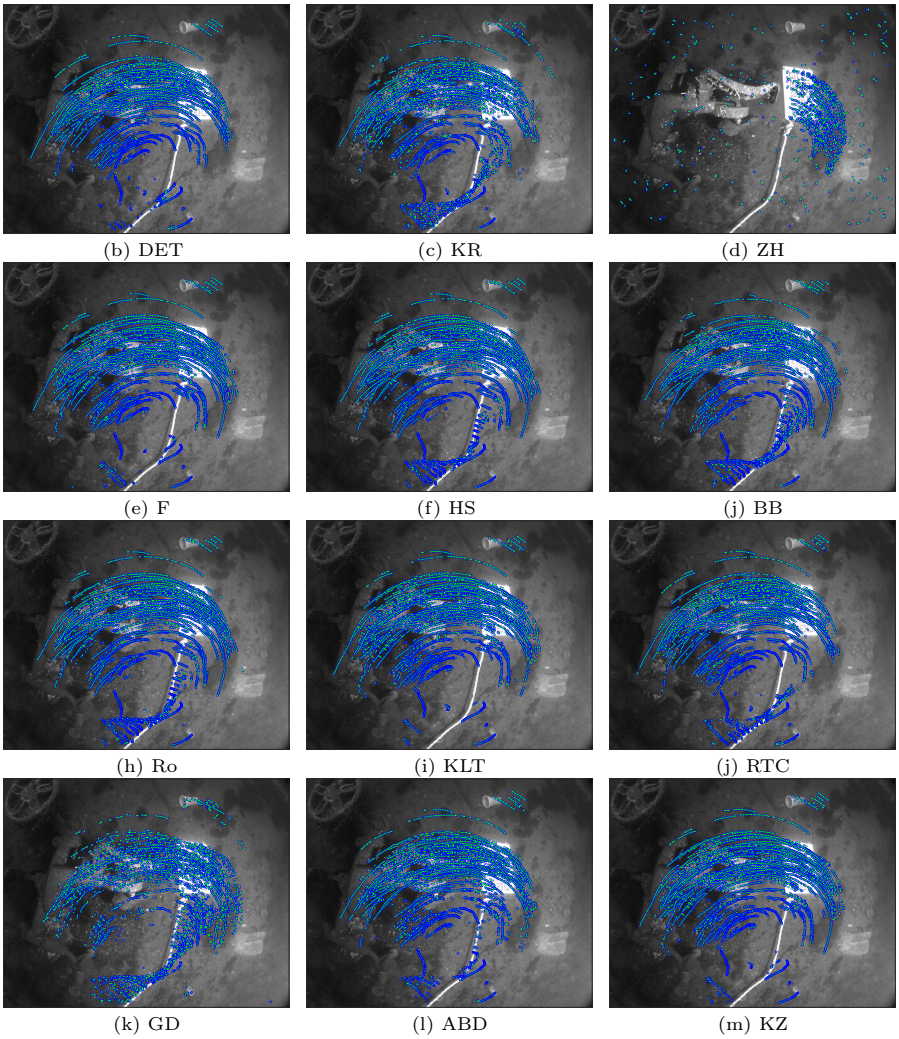
ing score, signifying higher stability of repeatability in key-point detection, HS (well-known Harris) as often called the benchmarking corner detector has not obtained the highest matching score among all, concerning the effects of image transformation detailed in Tab. 2. Furthermore, the detectors A, GD, KLT, and KZ have shown significant repeatability scores in the descending order. Besides the detectors, RTC, DET, KR and BB have shown low final scores $\mathcal{T}_S$ which their Std of matched points is also low. This objective repeatability assessment enables a valuable choice of the ρ scale parameter. Indeed, a bad ρ value regarding F, HS or Ro detectors (ρ=0.5 or 1, see Fig. 2) produces poor statistics than BB detector which is one of the least reliable detector (see Tab. 4). To conclude this part, since the movement of the camera is smooth, a score correlated to ZNCC constitutes a technique to assess the keypoint detectors –enabling to estimate the optimum ρ parameter–. Finally, the detector ZH repeatability score is the lowest, conveying that in the condition of numerous image transformations and noises (such as underwater videos), this detector is unreliable to use; and this drawback is mainly due to sensitivity of this detector to strong illumination changes.

To conclude this part, since the movement of the camera is smooth, a score correlated to ZNCC constitutes a technique to assess the repeatability of the keypoint detectors and enabling to estimate the optimum ρ parameter for several keypoint detectors.

Visual Results In Fig. 3(a), 50 random frame sequences have been chosen to illustrate the curves of the percentage of matched points. The 100 detected points along the sequence are displayed in blue in the first frame, while the linked points by the ZNCC are plotted in green in the Fig. 3(b)-(m). Visually the Ro, HS, and F are the best keypoint detection methods with smooth lines whereas the ZH, GD, KR and A detectors seem not stable because the displayed lines are sharped (zig-zag), illustrating the misplacement in the images of the Figs. 3 (c)-(d)-(k)-(l), reflecting the weak reliability of these techniques.



(a) Percentage of matched points per detector along 50 successive images



(b) DET (c) KR (d) ZH

(e) F (f) HS (j) BB

(h) Ro (i) KLT (j) RTC

(k) GD (l) ABD (m) KZ

Fig. 3. 100 matched keypoints on a sequence of 50 images. In (b)-(m), detected points appear in blue along the sequence, while links are in green.

4 Conclusion

In this paper, an extensive investigation of the 12 state-of-the-art keypoint detection techniques with the application to feature matching in the context of a complex video scene (AQUALOQ: containing all types of image transformation and various natural noises caused by the water and sand) has been conducted. The repeatability rate of each detection operator has been both statistically and visually demonstrated in presenting a guideline of which detectors are robust depending on video frame complexity. This work can contribute as a directive to the practitioners of this domain for choosing the appropriate keypoint detectors relating to the specific application (i.e., Ro, HS, and F exemplify the robust salient point detectors with the highest stability). The scale parameter ρ of the studied keypoint detectors have been studied for an objective and complete evaluation. The results show that KLT and Harris-Stephens (HS), two particularly popular detectors, perform well but not the best among the 12 tested, especially when the ρ parameter is not well selected.

This evaluation emphasizes on the filtering technique which is fast and straightforward than other approaches along with the keypoint matching methods (ZNCC) which is few time-consuming and can be easily implemented. Accordingly, the filtering techniques are useful for certain cases of image processing and optimization which are either used independently or can be used alongside with deep learning models either in pre-processing or post-processing stages.

In closing, this study could be performed on SLAM (Simultaneous Localization and Mapping) of these video sequences. SLAM is the most important problems in the pursuit of building truly autonomous mobile robots. With SLAM the spatial map of environment while simultaneously localizing the robot relative to this model can be acquired. The SLAM of this keypoint detection techniques will realize many new general purpose applications of mobile robotics.

References

1. Achard, C. and Bigorgne, E. and Devars, J.: A sub-pixel and multispectral corner detector. In: ICPR, pp. 959-962. (2000)
2. Awrangjeb, M., Guojun, L., Clive, F.: Performance Comparisons of Contour-Based Corner Detectors. IEEE TIP, **21**(9), 4167-4179 (2012)
3. Beaudet, P.R.: Rotationally invariant image operators, ICPR 1987, pp. 579-583.
4. Blom, J., Haar Romeny, B.M., Koenderink, J.J.: Affine invariant corner detection. Doctoral Dissertation, (1992)
5. Ferrera, M. and Creuze, V. and Moras, J. and Trouv, P. P.: AQUALOC: An Underwater Dataset for Visual-Inertial-Pressure Localization, The International Journal of Robotics Research, (2019)
6. Forstner, W. and Gulch, E.: A fast operator for detection and precise location of distinct points, corners and circular features, Fast Processing of Photogrammetric Data 1987, pp. 281-305.
7. Haggui, O., Tadonki, C., Lacassagne, L., Sayadi, F., Bouraoui, O.: Harris corner detection on a NUMA manycore. Future Generation Computer Systems. Elsevier, **88**, 442–452 (2018)

8. Harris, C. G and Stephens, M. J.: A Combined Corner and Edge Detector. In: Alvey Vision Conference, pp. 147-151. (1988)
9. Kenney, C.S. and Zuliani, M. and Manjunath, B.S.: An axiomatic approach to corner detection. In: CVPR, Vol. 1, pp. 191-197. (2005)
10. Kitchen, L.J., Rosenfeld, A.: Gray-level corner detection. Pattern Recognition Letters **1**, 95-102 (1982)
11. Lipschutz, M.-M.: Differential Geometry. McGfaw-Hill, (1969)
12. Mokhtarian, F., Mohanna, F.: Performance evaluation of corner detectors using consistency and accuracy measures **102**(1), 81-94 (2006)
13. Noble, J. Alison.: Finding corners. Image Vis. Comput. **6**(2), 121–128 (1988)
14. Rey-Otero, I. and Delbracio, M. and Morel, J. M.: Comparing feature detectors: A bias in the repeatability criteria, ICIP, pp. 3024-3028. (2015)
15. Rodehorst, V., Koschan, A.: Comparison and evaluation of feature point detectors. In: International Symposium Turkish-German Joint Geodetic Days, (2006)
16. Rohr, K.: Localization properties of direct corner detectors. Mathematical Imaging and Vision **4**, 139–150 (1994)
17. Rosten, E., Porter, R.,Drummond, T.: Faster and better: A machine learning approach to corner detection. TPAMI, IEEE, **32**(1), 105–119 (2008)
18. Schmid, C., Mohr, R., Bauckhange, C.: Evaluation of Interest Point Detectors. IJCV **37**, 151-172 (2010)
19. Shi, J. and Tomasi C.: Good features to track. In: CVPR, pp.593-600., (1994)
20. Shokouh, G.-S., Magnier, B., Xu, B., Montesinos, P. : Ridge Detection by Image Filtering Techniques: a Review and an Objective Analysis. PRIA **31**(2), 551–570 (2021)
21. Tuytelaars, T., Mikolajczyk, K.: Local Invariant Feature Detectors: A Survey. Now Foundations and Trends,(2008)
22. Wang, H. and Brady, M.: Real-time corner detection algorithm for motion estimation. IVC **13**(6), 695-703 (1995)
23. Xia, G., Delon, J., Gousseau, Y.: Accurate junction detection and characterization in natural images. IJCV. Springer, **106**(1), 31–56 (2014)
24. Zheng, Z., Wang, H., Teoh, E.-K.: Analysis of gray level corner detection. Pattern Recognit. Letters **20**(2), 149-162 (1999)
25. Canny, John.: A Computational Approach to Edge Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence **6**PAMI-8, 679-698 (1986)
26. Zuniga, O.-A., Haralick, R.-M.: Corner detection using the facet model. Pattern Recognit. and Image Processing, 30-37 (1983)