



## ARCHIVIO ISTITUZIONALE DELLA RICERCA

### Alma Mater Studiorum Università di Bologna Archivio istituzionale della ricerca

Keypoint detection by wave propagation

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

*Published Version:*

Keypoint detection by wave propagation / Salti, Samuele; Lanza, Alessandro; Di Stefano, Luigi. - In: JOURNAL OF ELECTRONIC IMAGING. - ISSN 1017-9909. - ELETTRONICO. - 30:1(2021), pp. 013003.1-013003.20. [10.1117/1.JEI.30.1.013003]

This version is available at: <https://hdl.handle.net/11585/790768> since: 2021-01-23

*Published:*

DOI: <http://doi.org/10.1117/1.JEI.30.1.013003>

*Terms of use:*

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

(Article begins on next page)

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).  
When citing, please refer to the published version.

This is the final peer-reviewed accepted manuscript of:

**Salti, S., Lanza, A., & Di Stefano, L. (2021). Keypoint detection by wave propagation. Journal of Electronic Imaging, 30(1).**

The final published version is available online at: <http://dx.doi.org/10.1117/1.JEI.30.1.013003>

Rights / License:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

*This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>)*

***When citing, please refer to the published version.***

# Keypoint Detection by Wave Propagation

Samuele Salti<sup>a</sup>, Alessandro Lanza<sup>b</sup>, Luigi Di Stefano<sup>a</sup>

<sup>a</sup>University of Bologna, Department of Computer Science and Engineering, Viale Risorgimento 2, 40135, Bologna, Italy

<sup>b</sup>University of Bologna, Department of Mathematics, Piazza di Porta San Donato 5, 40126, Bologna, Italy

**Abstract.** This paper proposes to rely on the wave equation for the detection of repeatable keypoints invariant up to image scale and rotation and robust to viewpoint variations, blur and lighting changes. The algorithm exploits the properties of local spatial-temporal extrema of the evolution of image intensities under the wave propagation to highlight salient symmetries at different scales. While the image structures found by most state-of-the-art detectors, such as blobs and corners, occur typically on highly textured surfaces, salient symmetries are widespread in diverse kinds of images, including those related to poorly textured objects, which are hardly dealt with by current pipelines based on local invariant features. The impact on the overall algorithm of different numerical wave simulation schemes and their parameters is discussed, and a pyramidal approximation to speed-up the simulation proposed and validated. Experiments on publicly available datasets show that the proposed algorithm offers state-of-the-art repeatability on a broad set of different images while detecting regions that can be distinctively described and robustly matched.

**Keywords:** Feature extraction, Finite difference methods, Wave equation.

\*Samuele Salti, [samuele.salti@unibo.it](mailto:samuele.salti@unibo.it)

## 1 Introduction

Partial Differential Equations (PDEs) are a common tool in the field of image processing. Their interpretation as Euler-Lagrange solutions of variational optimization problems provides an elegant unified framework for diverse but related basic problems such as denoising, deblurring and inpainting. A variety of PDEs are currently being used and provide state-of-the-art approaches for these problems. The most popular is certainly the heat or diffusion equation, but other non-linear second-order,<sup>1</sup> fourth-order<sup>2</sup> and, very recently, fractional-order PDEs<sup>3</sup> have been deployed for image processing. Although the great majority of these equations are parabolic, researchers are also investigating on the use of hyperbolic equations, such as shock-filters<sup>4</sup> or the telegrapher equation.<sup>5</sup>

PDEs are notably less investigated in image analysis and computer vision problems, one prominent exception being the linear scale–space based on the heat equation deployed for scale-invariant

keypoint detection. Invariant keypoint detection and matching is a fundamental problem in Computer Vision, with applications in image retrieval, object detection and recognition, object and camera tracking, camera calibration, and image registration, among the others. In this paradigm, the image content is summarized by a set of small, possibly overlapping representative regions, which are repeatably found at points corresponding to the same world points regardless of distortions due to changes in the acquisition process, the illumination or the viewpoint. Corresponding features can then be identified by looking for the nearest neighbor in the space of descriptions of the regions, providing both a sparser and more robust representation of the image content.

The scale–space theory was developed in seminal works by Witkin<sup>6</sup> and Koenderink,<sup>7</sup> and later popularized by the automatic scale selection principle proposed by Lindeberg.<sup>8</sup> Lindeberg’s scale–normalized derivatives provide the theoretical foundations for the Differences-of-Gaussians (DoG) keypoint detector,<sup>9</sup> arguably among the most influential contribution in the field of computer vision. However, other kinds of PDEs are unexplored in problems related to detection and/or description of local invariant features. Instead, in work dating back to the 90’s,<sup>10–12</sup> the wave equation was used for skeletonization of binary silhouettes and detection of circular objects, due to its ability of eliciting the symmetry set. More recently, PDEs other than the heat equation have been used for description of global silhouettes.<sup>13</sup>

In this paper we investigate the use of other PDEs than the heat equation in image analysis from the modern perspective of local invariant features. In particular, we focus on the problem of invariant keypoints detection and advocate the use of the wave equation. When applied to images, the wave equation exhibits the ability to highlight symmetries at different scales. Salient symmetries are likely to be found in a large range of images, and in particular in those related to man-made untextured objects, where the established recognition paradigm based on matching local

features is prone to fail due to scarcity of cues, especially when occlusions need to be withstood. Nevertheless, many salient symmetries arise in textured objects alike. Thus, the proposed approach qualifies as a general tool for repeatable salient regions detection. A preliminary version of this study was published in.<sup>14</sup>

## 2 Related work

There is a rich literature on the topic of keypoint detection, as well as several surveys and evaluations such as,<sup>15-18</sup> and a review on this broad topic is beyond the scope of this paper. We briefly cite here the main proposals and refer the interested reader to the above mentioned surveys.

The first proposals that can be defined as interest point detectors are the Morevec and Harris corner detectors,<sup>19,20</sup> invariant up to rotations. Scale invariance by automatic scale selection was introduced and deeply studied by Lindberg,<sup>8</sup> and his ideas fostered the development of several scale-invariant detectors, such as the Difference of Gaussians<sup>9</sup> and the FastHessian blob detectors.<sup>21</sup> The Harris corner detector was later extended to scale and affine invariance by Mikolajczyk and Schmid.<sup>22</sup> A widely used affine invariant region detector is the Maximally Stable Extremal Regions algorithm (MSER) by Matas *et al.*<sup>23</sup> A family of corner detectors focused on efficiency are based on the Accelerated Segment Test (AST) by Rosten and Drummond,<sup>24</sup> *i.e.* a set of intensity tests to be performed on a Bresenham's circle. This family includes the machine-learning based FAST detector in,<sup>25</sup> where the order of the tests is learned in a tree from a training set, later improved in;<sup>26</sup> the AGAST detector,<sup>27</sup> where two trees are used to classify separately homogeneous and textured regions; the BRISK<sup>28</sup> and ORB<sup>29</sup> detectors, which endow the FAST approach with scale and rotation invariance, respectively.

More recent proposals include: the SFOP detector,<sup>30</sup> which proposes a scale-selection mecha-

nism for junction-type features; the SIFER detector,<sup>31</sup> which uses the Cosine Modulated Gaussian filter and trades rotation invariance for higher scale localization accuracy and resilience to blur and lighting changes; KAZE features<sup>32</sup> and their faster extension Accelerated KAZE,<sup>33</sup> which create the scale-space by using non-linear diffusion filters to make blurring locally adaptive and retain object boundaries.

Surprisingly, there has been relatively little work on the use of symmetry as a cue to detect and describe local features. The most recent contributions on this topic are due to Maver<sup>34</sup> and Haugge et Snavely.<sup>35</sup> In particular, the latter paper proposes a detector-descriptor pair relying on a local symmetry score computed densely on the image and across scales. Unlike our proposal, though, their formalization of symmetry is specifically engineered to capture effectively the salient regions likely to be found in architectural images and relies on the classic Gaussian scale-space rather than our novel formulation grounded on the wave PDE. Earlier related works such as<sup>36,37</sup> focus on detecting interest points featuring radial symmetries (*e.g.* eye centres) but are neither really conceived nor evaluated as local invariant features for general purpose image matching tasks.

### **3 Wave-based Detection**

In this section, we illustrate in detail the proposed keypoint detector. In particular, in Sec. 3.1 we present the continuous PDE model based on the wave equation and in Sec. 3.2 we show the scale covariant properties of such model. Sec. 3.1 and 3.4 are dedicated to the numerical solution of the proposed hyperbolic initial boundary-value problem. In particular, Sec. 3.1 presents a class of compact (explicit and implicit) finite difference schemes, which efficiently compute a numerical solution of the problem; in Sec. 3.4 we analyze stability and numerical dispersion of the considered schemes, with particular attention to isotropicity of the schemes which is crucial for rotation

invariance of the detector. Finally, in Sec. 3.5 we illustrate the procedure used for detecting and validating the spatio-temporal extrema of the computed numerical solution.

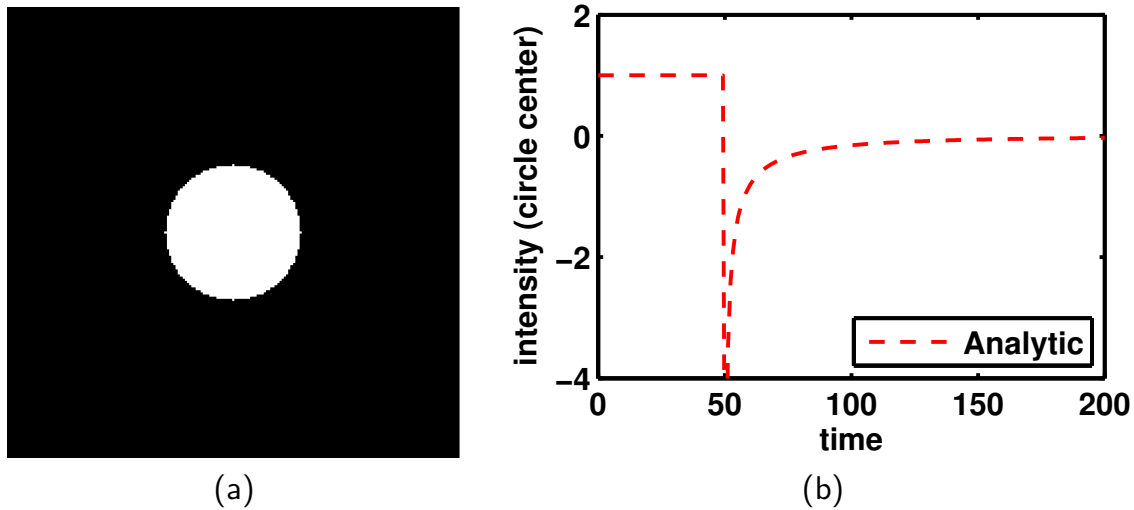
### 3.1 The wave equation

The wave equation is a well-known linear second-order hyperbolic partial differential equation. It is used in several fields of physics that need to model waves propagating through time and space, such as in acoustics, electromagnetism and fluid dynamics.

The initial boundary value problem modeling wave propagation that we use is as follows:

$$\left\{ \begin{array}{ll} u_{tt}(x, y; t) = c^2 \nabla^2 u(x, y; t) & (x, y; t) \in \Omega \times \mathbb{R}^+ \\ u(x, y; 0) = I(x, y) & (x, y) \in \bar{\Omega} \\ u_t(x, y; 0) = 0 & (x, y) \in \bar{\Omega} \\ u_t(x, y; t) = c u_n(x, y; t) & (x, y; t) \in \partial\Omega \times \mathbb{R}^+ \end{array} \right. \quad (1)$$

where  $c \in \mathbb{R}^+$  is the wave speed,  $I$  is the image,  $\bar{\Omega} = (\Omega \cup \partial\Omega) \subset \mathbb{R}^2$  is the (rectangular) image domain,  $n$  is the inward pointing unit normal to the image domain boundary  $\partial\Omega$ . The solution  $u(x, y; t)$  models the height of the wave, *i.e.* the graylevel intensities at each space-time location.  $\nabla^2$  is its 2-D spatial Laplacian operator,  $u_t$  and  $u_{tt}$  are its first-order and second-order partial derivatives along the time dimension, while  $u_n(x, y; t)$  is the partial derivative in space along direction  $n$ . The initial conditions, defined by the second and third equations, are trivial. The boundary conditions, defined by the fourth equation, represent the first order highly absorbing local approximation of the perfectly absorbing boundary conditions,<sup>38</sup> which are inherently non-local in space and time, *i.e.* to advance one time level at a single point they require information over



**Fig 1** (a) initial image  $I$ ; (b) analytical solution of the wave equation at the circle center.

the entire boundary from all previous times. Absorbing boundary conditions have been chosen to avoid reflections of waves at image boundaries, that would cause unwanted interferences. The selected initial and boundary conditions render problem (1) well-posed.<sup>38</sup>

### 3.2 Scale covariant analysis

The stack of images obtained by letting the image intensities evolve under the process simulated by the wave equation (1) enables a multi-scale analysis of the input image. In particular, as shown in,<sup>10</sup> by simulating wave propagation on images it is possible to detect circles of varying radii. This derives directly from Huygens' principle of superposition: if we consider a circle as depicted in Fig. 1, and interpret the intensities as the wave height at time 0, then the wave front after the wave has traveled a distance equal to the radius of the circle is given by the sum of circular waves originating from each point on the circle edge. Therefore, the wave propagation attains an extremum at the center of the circle at a time that is proportional to the scale of the original circle.

This is confirmed by the analytical solution of the equation when  $I(x, y)$  represents a uniform



circle of radius  $r$  on a uniform background, as depicted in Fig. 1. In such a case, letting  $k_1$  and  $k_2$  denote, respectively, the greylevel inside and outside the circle, the solution  $u(x, y; t)$  restricted to the circle center  $(x_0, y_0)$ , is given by

$$u(x_0, y_0; t) = \begin{cases} k_1 & t \leq \frac{r}{c} \\ k_1 + \frac{(k_2 - k_1)}{\sqrt{1 - (\frac{r}{ct})^2}} & t > \frac{r}{c} \end{cases} \quad (2)$$

The plot in Fig. 1 (b) shows that the function exhibits a vertical asymptote in the continuous domain at a time proportional to the radius of the circle: having set the wave velocity equal to 1, the asymptote shows up after 50 time intervals, which is exactly the radius in pixels of the circle.

Detection of circle centers is actually a particular case of the more general property of wave propagation of eliciting the symmetry set of curves.<sup>12</sup> The symmetry set is defined as the locus of centers of circles bi-tangent to a curve and can be detected by summing  $u(x, y; t)$  over  $t$  into an accumulator and inserting the local spatial extrema of the accumulator into the set at each iteration. Such an approach detects all kind of symmetries, regardless of their saliency, and it is effective to skeletonize binary shapes. Differently, to define repeatable keypoints, we propose to consider only the points of locally maximal symmetry, which correspond to the shock points of the wave propagation (see<sup>13</sup> for a deep analysis of shock points). They can be detected by analyzing the solution  $u(x, y; t)$  of (1).

Given the property of  $u(x, y; t)$  to enable scale-covariant analysis of images, and slightly abusing the terminology, we may think of it as a *wave-based scale-space*, although this one-parameter family of signals does not satisfy the classical scale-space axioms.<sup>8</sup>

### 3.3 Numerical solutions of the wave equation

To solve (1) by numerical simulation several finite difference schemes, both explicit and implicit, may be deployed. Here we restrict ourselves to describe and evaluate only compact schemes, *i.e.* schemes with stencils not wider than 3 nodes in the discrete space-time domain. By defining the regular space-time domain mesh  $y_i = i \Delta y$ ,  $x_j = j \Delta x$ ,  $t^n = n \Delta t$  and using the classical centered finite difference operators

$$\delta_x^2 u_{i,j}^n \equiv u_{i,j+1}^n - 2u_{i,j}^n + u_{i,j-1}^n \quad (3)$$

$$\delta_y^2 u_{i,j}^n \equiv u_{i+1,j}^n - 2u_{i,j}^n + u_{i-1,j}^n \quad (4)$$

$$\delta_t^2 u_{i,j}^n \equiv u_{i,j}^{n+1} - 2u_{i,j}^n + u_{i,j}^{n-1} \quad (5)$$

to approximate second order derivatives, where  $u_{i,j}^n \simeq u(x_j, y_i, t^n)$ , a family of compact schemes approximating the solution can be defined as follows:<sup>39</sup>

$$\left[1 + a (\delta_x^2 + \delta_y^2) + a^2 \delta_x^2 \delta_y^2\right] \delta_t^2 u_{i,j}^n = \lambda^2 \left[(\delta_x^2 + \delta_y^2) + b \delta_x^2 \delta_y^2\right] \delta_t^2 u_{i,j}^n \quad (6)$$

where  $\lambda = c\Delta t/\Delta x \in \mathbb{R}^+$  is the Courant number, and  $a \in [-\frac{1}{4}; \frac{1}{4}]$ ,  $b \in [0; \frac{1}{2}]$  are numerical parameters. The resulting scheme is explicit for  $a = 0$ , implicit otherwise.

For implicit schemes, (6) does not represent the most general family of solutions, but the family of those that can be solved efficiently with the Alternating Direction Implicit (ADI) method,<sup>40</sup> which allows to use the Thomas algorithm for efficient matrix inversion, whose complexity is  $O(N)$  instead of  $O(N^3)$ . The Thomas algorithm works with tridiagonal matrices. To obtain such matrices, the general formulation is split into two equations, which are solved sequentially as

follows:

$$(1 + a\delta_x^2) u_{i,j}^{n+1*} = \frac{\lambda^2}{a} [-1 + (a - b) \delta_y^2] u_{i,j}^n \quad (7)$$

$$(1 + a\delta_y^2) \delta_t^2 u_{i,j}^n = u_{i,j}^{n+1*} + \frac{\lambda^2}{a} (1 + b\delta_y^2) u_{i,j}^n . \quad (8)$$

By varying  $a$  and  $b$ , different implicit schemes are obtained. According to the analysis presented in,<sup>41</sup> notable special cases can be singled out. In the following, we restrict ourselves to the values of  $a$  and  $b$  that result in the so called Maximally Flat Isotropic (MFI) scheme, as it is that for which the difference between axial and diagonal relative phase velocity is maximally flat, and should therefore offer the best rotation invariance. Values of  $a$  and  $b$  for MFI and the schemes that will be presented below are reported in Tab. 1.

For explicit schemes, the general solution becomes:

$$\delta_t^2 u_{i,j}^n = \lambda^2 [(\delta_x^2 + \delta_y^2) + b\delta_x^2\delta_y^2] \delta_t^2 u_{i,j}^n \quad (9)$$

which yields the following time-marching scheme:

$$u_{i,j}^{n+1} = \lambda^2 \begin{bmatrix} b & 1 - 2b & b \\ 1 - 2b & 4b - 2 & 1 - 2b \\ b & 1 - 2b & b \end{bmatrix} \otimes u_{i,j}^n - u_{i,j}^{n-1} \quad (10)$$

where  $\otimes$  represents convolution. Explicit solutions correspond to different approximations of the Laplacian operator. Hence, for explicit schemes,  $b$  controls the interpolation between the two

extreme of the spectrum of explicit solutions, where only four of the eight neighbors of a node are actually used: the Standard Leapfrog scheme (SLF), obtained with  $b = 0$ , which uses horizontal and vertical neighbors only, and the Rotated Leapfrog (RLF) scheme, obtained with  $b = \frac{1}{2}$ , which uses the neighbors along the diagonals. Here we consider the two notable interpolation highlighted in,<sup>41</sup> namely the INT 1/4 and INT 1/6 schemes, which correspond to setting  $b = 1/4$  and  $b = 1/6$ , respectively, as they are the most promising as far as rotation invariance is concerned.

The most costly operation in the update equation (10) is convolution. It can be made less demanding by modifying the filter to be linearly separable, so that it requires two convolutions with  $3 \times 1$  kernels only. Though the  $3 \times 3$  kernel is not directly separable, it becomes such as we move part of the contribution of the central pixel outside of the convolution. One can readily verify that the following equation

$$u_{i,j}^{n+1} = \begin{bmatrix} \lambda\sqrt{b} \\ \lambda\frac{1-2b}{\sqrt{b}} \\ \lambda\sqrt{b} \end{bmatrix} \otimes \left( \begin{bmatrix} \lambda\sqrt{b} & \lambda\frac{1-2b}{\sqrt{b}} & \lambda\sqrt{b} \end{bmatrix} \otimes u_{i,j}^n \right) + \left( 2 - \frac{\lambda^2}{b} \right) u_{i,j}^n - u_{i,j}^{n-1} \quad (11)$$

is equivalent to (10).

The implicit and explicit schemes provide the numerical approximations  $u_{i,j}^n \simeq u(x_i, y_j, t^n)$  of the solution at the *inner* mesh nodes, *i.e.* for  $(i, j, n) \in \{2, \dots, w-1\} \times \{2, \dots, h-1\} \times \{2, \dots\}$ . To compute the solution at the inner nodes for the first iteration, we exploit the initial condition on the time partial derivative, whose discretization with central differences yields  $u_{i,j}^1 - u_{i,j}^{-1} = 0$ .

**Table 1** Numerical schemes parameters for wave simulation.  $\lambda_S$  and  $\lambda_{NS}$  refer to the Courant number used to test schemes on the stability bound (S) and in the middle of their stability range (NS).

Scheme	a	b	Stability	$\lambda_S$	$\lambda_{NS}$
INT (1/4)	0	$\frac{1}{4}$	$\lambda \leq 1$	1	$\frac{1}{2}$
INT (1/6)	0	$\frac{1}{6}$	$\lambda \leq \frac{\sqrt{3}}{2}$	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$
MFI	$\frac{1}{4} - \frac{1}{2\sqrt{3}}$	$\frac{1}{6}$	$\lambda \leq 1$	1	$\frac{1}{2}$

This, in turn, leads to the following solution for the implicit case

$$(1 + a\delta_x^2) u_{i,j}^{1*} = \frac{\lambda^2}{a} [-1 + (a - b)\delta_y^2] u_{i,j}^0 \quad (12)$$

$$(1 + 2a\delta_y^2) u_{i,j}^1 = u_{i,j}^{1*} + \left(1 + \frac{\lambda^2}{a} + \left(a + \frac{b\lambda^2}{a}\right)\delta_y^2\right) u_{i,j}^0 \quad (13)$$

and the explicit case

$$u_{i,j}^1 = \frac{1}{2} \begin{pmatrix} \lambda\sqrt{b} \\ \lambda\frac{1-2b}{\sqrt{b}} \\ \lambda\sqrt{b} \end{pmatrix} \otimes \left( \begin{bmatrix} \lambda\sqrt{b} & \lambda\frac{1-2b}{\sqrt{b}} & \lambda\sqrt{b} \end{bmatrix} \otimes u_{i,j}^0 \right) + \left(2 - \frac{\lambda^2}{b}\right) u_{i,j}^0. \quad (14)$$

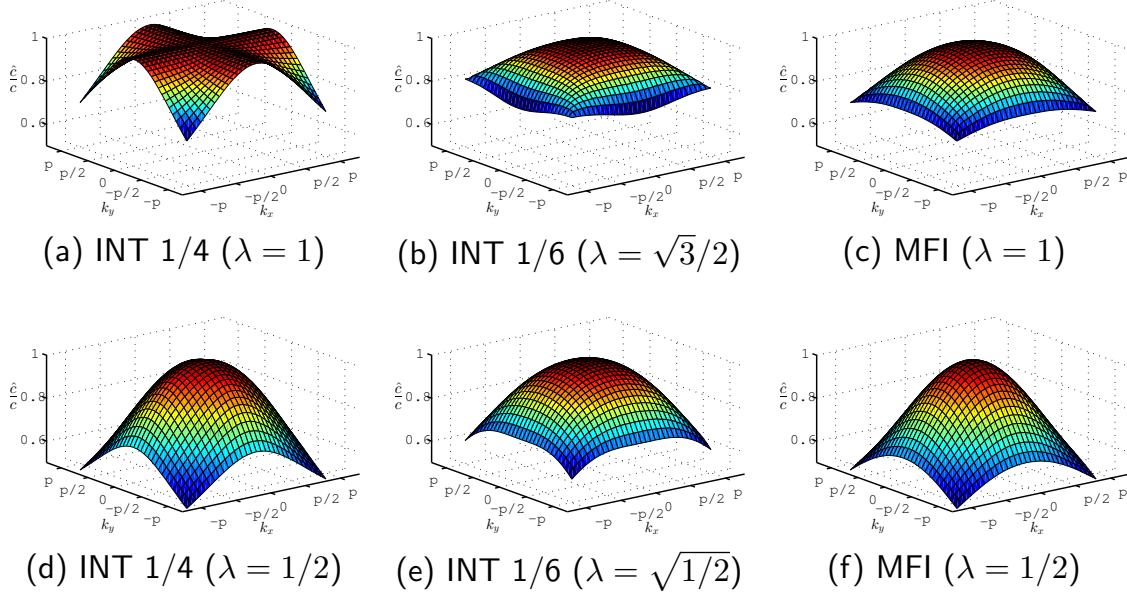
As for the boundary nodes, we exploit the boundary conditions, whose discretization with forward differences along the time and normal-to-the-boundary axis yields the following solution for both kinds of numerical schemes:

$$u_{i,1}^{n+1} = u_{i,1}^n + \lambda(u_{i,2}^n - u_{i,1}^n) \quad (15)$$

$$u_{1,j}^{n+1} = u_{1,j}^n + \lambda(u_{2,j}^n - u_{1,j}^n) \quad (16)$$

$$u_{i,w}^{n+1} = u_{i,w}^n + \lambda(u_{i,w-1}^n - u_{i,w}^n) \quad (17)$$

$$u_{h,j}^{n+1} = u_{h,j}^n + \lambda(u_{h-1,j}^n - u_{h,j}^n). \quad (18)$$



**Fig 2** Numerical dispersion. First row: INT 1/4, INT 1/6 and MFI with Courant number on the stability bound. Second row: INT 1/4, INT 1/6 and MFI with smaller Courant number.

### 3.4 Numerical stability and dispersion

Numerical stability of the reviewed schemes depends on the Courant number  $\lambda$ . As, in the case of images,  $\Delta x = \Delta y$  is given by the pixel side, the stability constraint actually affects the product  $c\Delta t$  *i.e.* the space traveled by the wave in one unit of discretized time, rather than the two constants individually. Different schemes have different bounds on this parameter, as reported in Tab. 1. To vary the Courant number within its stability bounds we opted to keep the velocity  $c$  constant to  $\frac{\sqrt{2}}{2}$  and vary  $\Delta t$  accordingly.

Unfortunately, discretization of the wave propagation is prone to quite significant numerical errors, mainly due to numerical (or grid) dispersion.<sup>42</sup> Numerical dispersion stems from the numerical approximation of derivatives and operators, which in turn lets different frequencies in the original image travel with different actual speeds during the numerical simulation. Therefore, a useful measure of numerical dispersion in wave simulations is the ratio of the effective numerical

wave speed  $\hat{c}$  over the real wave speed  $c$ , *i.e.* the relative phase velocity. As it can be observed in Fig. 2, different numerical schemes let the wave travel with different relative phase velocities along different directions, and dispersion is also affected by the Courant number resulting from the choice of  $\Delta t$ . As numerical dispersion cannot be eliminated, the preferable dispersion in our context is that distorting the original signal isotropically along all directions, so that rotation invariance of the detector is mostly preserved. We can clearly see that by using a Courant number not at the stability bound, all the schemes become more isotropic than their counterpart at the stability bound. This is even more evident for the INT 1/4 scheme, which we originally used in the preliminary version of this work.<sup>14</sup> The higher isotropy comes at the cost of a lower relative phase velocity. The INT 1/6 scheme exhibits the smallest rise in numerical dispersion when the schemes are not run at the stability bound.

For every scheme, there are some frequencies that are not well propagated, *i.e.* whose relative phase velocity is low. Their number is even higher in isotropic schemes. In particular, for every scheme we can compute a cut-off frequency for each direction, above which waves become evanescent and propagation is not correct. This can be clearly seen in Fig. 3, left, where the different propagation speeds of the frequencies of the circle image (Fig. 1 (a)) cause spurious oscillations at the circle center over time, resulting in a very different response compared to the analytic solution depicted in Fig. 1 (b). To overcome such issue, we adopt a solution similar to that proposed in:<sup>12</sup> we interleave a wave propagation step and a linear diffusion step governed by the heat equation with diffusivity  $k \in \mathbb{R}^+$ ,

$$u_t(x, y; t) = k \nabla^2 u(x, y; t), \quad (19)$$

which we implement using the discrete spatial Laplacian approximation used in<sup>12</sup> and forward differences for the time derivative. These choices result in the following explicit scheme, that we interleave with a step of wave propagation,

$$u_{i,j}^{n+1} = \begin{bmatrix} p & 2p & p \\ 2p & 1 - 12p & 2p \\ p & 2p & p \end{bmatrix} \otimes u_{i,j}^{n+\frac{1}{2}} \quad (20)$$

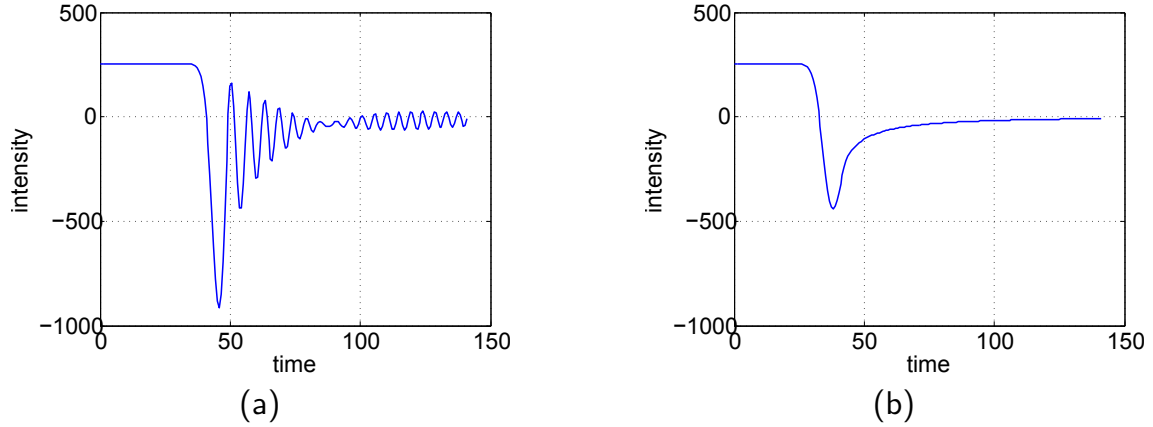
where  $p = \frac{k\Delta t}{\Delta x^2}$  is a constant and  $u^{n+\frac{1}{2}}$  is the output of the last wave propagation step.

We use Neumann boundary conditions for the heat equation, *i.e.*

$$\begin{aligned} u_{i,1}^{n+1} &= u_{i,2}^{n+1} & u_{1,j}^{n+1} &= u_{2,j}^{n+1} \\ u_{i,w}^{n+1} &= u_{i,w-1}^{n+1} & u_{h,j}^{n+1} &= u_{h-1,j}^{n+1}. \end{aligned} \quad (21)$$

The beneficial effect of the wave-diffusion process is clearly visible in Fig. 3, right, where spurious oscillations have been filtered out. Unfortunately, the diffusion process also negatively affects the numerical solution: it reduces the strength of the extrema, thereby making it harder to robustly detect them, and it makes them appear earlier, thus reducing the theoretically perfect scale covariance of the wave solution. In spite of these drawbacks, in what follows we make always use of the interleaved process to remove spurious oscillations and enable robust extrema detection.





**Fig 3** Temporal evolution of the intensity at the center of the circle in Fig. 1 under the discretized wave process (a) and the discretized mixed wave-diffusion process (b).

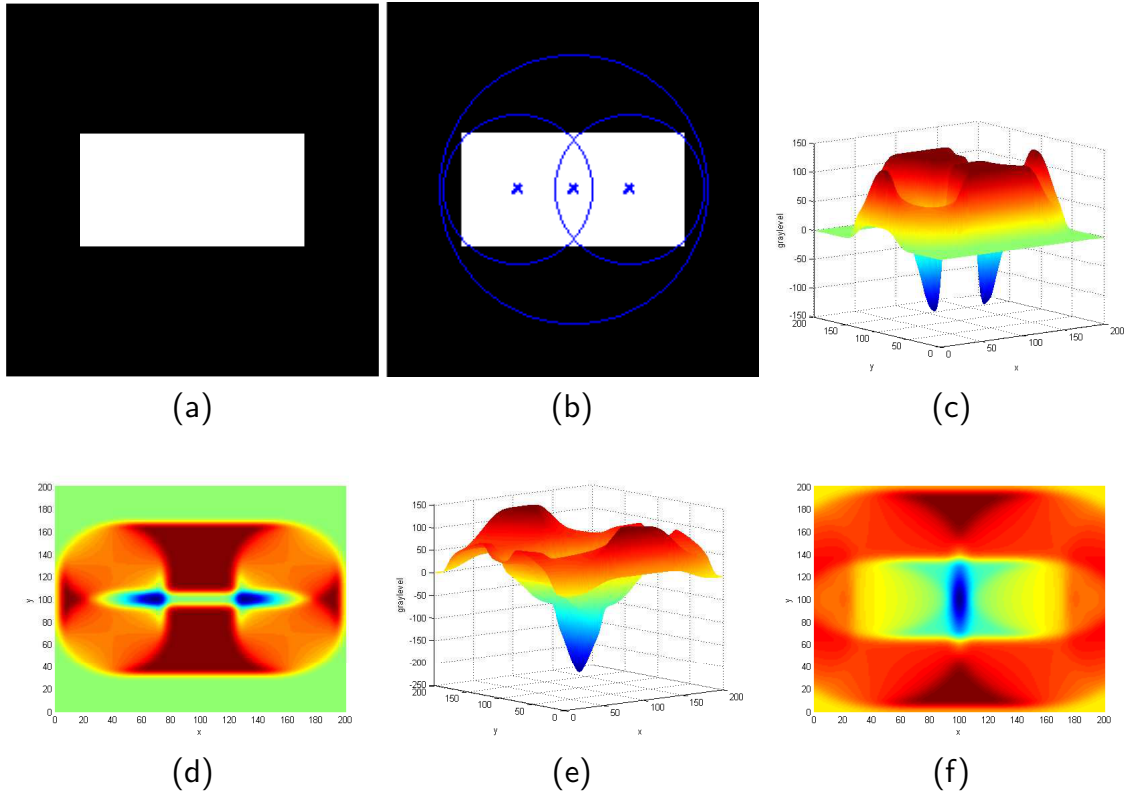
### 3.5 Extrema detection and validation

As illustrated in Fig. 4, the spatio-temporal extrema of  $u(x, y; t)$  identify perceptually meaningful image structures: in the considered example, three extrema show up, two corresponding to the symmetries arising at the center of the square-like structures formed by the sides of the rectangle and one corresponding to the center of the rectangle.

The time of a spatio-temporal extremum  $(x_{ext}, y_{ext}, t_{ext})$  is related to the scale (the radius)  $r_{ext}$  of the detected symmetry by the following simple relation:

$$r_{ext} = \frac{c t_{ext}}{\Delta x} = \frac{c \Delta t n_{ext}}{\Delta x}. \quad (22)$$

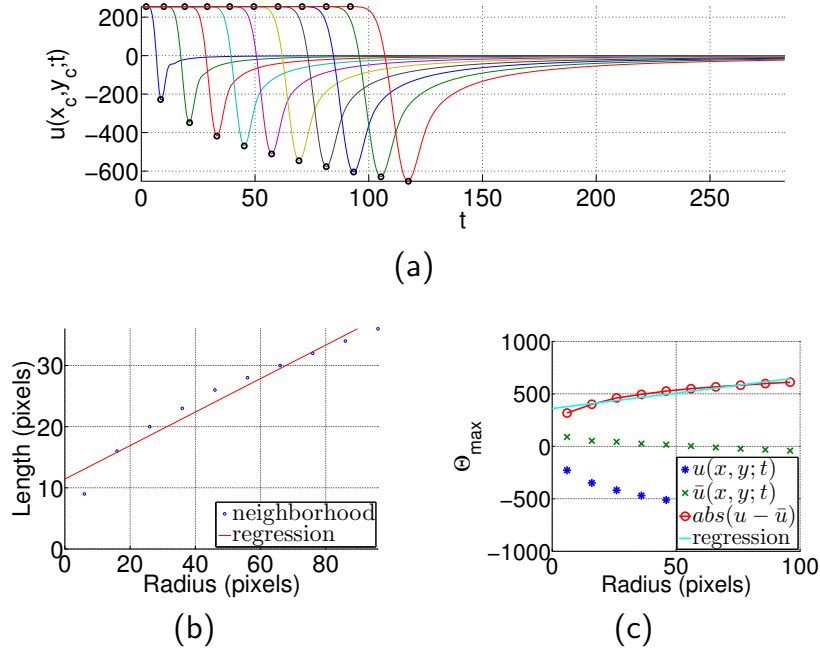
Such scale is directly expressed in pixels. This is a relevant difference in practice with respect to diffusive scale-spaces, where the relationship between the scale  $\sigma$  and the associated image patch is somewhat arbitrary, due to the theoretically infinite extent of the Gaussian kernel.



**Fig 4** (a) initial image  $I$ ; (b) extrema of  $u(x, y; t)$ ; (c-d) wave state when the two smaller keypoints are detected, from a side and a top view, respectively; (e-f) wave state when the third keypoint is detected.

Given such scale and rotation covariant extrema in our family of signals, we define as keypoints the *sharp* local extrema of  $u(x, y; t)$ . Although in the synthetic images considered so far the requirement of sharpness of local extrema is redundant, as all extrema are indeed sharp, in real images, weak (*i.e.* not repeatable) symmetries may be detected if all local extrema are accepted rather than *sharp* ones only.

We define an extremum as sharp if its value is significantly larger (or smaller) than the average of the values in its temporal neighborhood:



**Fig 5** Definition of sharp extrema: (a) Temporal evolution of the intensity at the center  $(x_c, y_c)$  of circles of varying radius. The extrema (minima) show up at times proportional to the circles radius. Black dots on each curve indicate the neighborhood used at each scale for sharpness-based keypoints filtering; (b) size of the neighborhoods defined by the black dots in (a) and their linear regression as a function of radius; (c) linear regression as a function of radius of the absolute difference between the value at the extremum and the mean value in the neighborhood .

$$|u(x_{ext}, y_{ext}, t_{ext}) - \bar{u}(x_{ext}, y_{ext}, t_{ext})| \geq \theta, \quad (23)$$

$$\bar{u}(x_{ext}, y_{ext}, t_{ext}) = \frac{1}{t_2 - t_1} \sum_{k=t_1}^{t_2} u(x_{ext}, y_{ext}, k). \quad (24)$$

Both the neighborhood  $[t_1, t_2]$  and the threshold  $\theta$  must adapt with scale, as the absolute value at the extrema as well as their smoothness increase with scale (Fig. 5 (a)). They also change according to the contrast at the edges defining the symmetry. Sharpness could be measured by considering spatial and/or temporal neighborhoods. However, spatial neighborhoods are not defined at the boundary of the image and the computation of the average is more expensive than over

1-D temporal neighborhoods (efficient schemes, such as integral images or histograms can be, of course, deployed in both cases). We experimentally found that measuring sharpness only along time suffices to prune unwanted weak symmetries.

The definition of the neighborhood must fulfill two contradicting requirements: it should be small enough to avoid the influence of clutter and occlusions in real applications, but also large enough to allow for discriminating between weak and repeatable symmetries. To limit the influence of clutter and occlusions, we select as end of the neighborhood the extremum itself,  $t_2 = t_{ext}$ , so that any texture external to the symmetry does not influence its detection. To define the starting point of the neighborhood,  $t_1$ , we analyze the behavior of the most symmetrical shape with the highest contrast, *i.e.* a circle like that depicted in Fig. 1. The trend of  $u(x, y; t)$  at the center of the circle starts with a plateau and then exhibits a sudden drop toward the extremum: indeed, this would be a discontinuity in the continuous case (Fig. 1). Therefore, we define the starting point of the neighborhood as the end of the plateau<sup>1</sup> (Fig. 5 (a)).

As the difference between the discrete solution and the continuous solution grows when the scale gets larger due to both numerical dispersion and the diffusion interleaved with propagation, we learn the length  $t_{21} = |t_2 - t_1|$  of the time interval to compute the sharpness for each scale by linear regression of the model  $t_{12}(r_{ext}) = a_t r_{ext} + b_t$  (Fig. 5 (b)). Using again the image of the circle, we also regress a linear model  $\theta_{max}(r_{ext}) = a_\theta r_{ext} + b_\theta$  for the value of the absolute difference  $|u - \bar{u}|$  (Fig. 5 (c)): this defines the maximum value  $\theta_{max}$  for the threshold  $\theta$ . We would like to point out that these regressions are performed only once to estimate the parameters, which are then used to detect keypoints on every kind of input image. As we are interested in detecting

---

<sup>1</sup>The end of the plateau is identified as the first point  $n_s$  along the time axis where the absolute value of the time derivative is larger than a threshold value, *i.e.*  $\text{abs}(u_{i,j}^{n_s+1} - u_{i,j}^{n_s}) > th$ . We choose  $th = 1.5$  in our regressions.

weaker, though repeatable, symmetries than a high-contrast circle, the actual value  $\theta$  used at each scale is defined as a fraction  $\rho$  of  $\theta_{max}(r_{ext})$ .

#### 4 Simulation with pyramid of images

One of the main limits of resorting to wave propagation to extract keypoints is the computational cost of the simulation. To overcome such limit, we propose to deploy an image pyramid, as done *e.g.* in.<sup>9</sup> As in the standard image pyramid, we consider a fixed number of images per octave, and we down-sample each side of the image by a factor of 2 after each octave.

When down-sampling the image to move to a different octave, we keep the Courant number constant. This results in a scheme that is still stable, but simulates waves by using a larger  $\Delta t$ . Indeed, at octave  $i$ , the corresponding  $\Delta t^{(i)}$  is

$$\Delta x^{(i)} = 2^i \Delta x, \quad \lambda^{(i)} = \lambda^{(0)} \Rightarrow \quad \Delta t^{(i)} = 2^i \Delta t \quad (25)$$

Given the radius  $r_{max}$  of the biggest feature we wish to detect and the number  $L$  of the images per octave, we compute the number of octaves  $O$  by the following formula:

$$O = \underset{o}{\operatorname{argmin}} \sum_{i=0}^o L * 2^i \leq T_{max} = \left\lceil \frac{r_{max}}{c\Delta t} \right\rceil \quad (26)$$

The advantage of simulating wave propagation by a pyramidal scheme is twofold: the cost to simulate one time step of wave propagation decreases due to the smaller number of pixels at the higher octaves; the cost to reach the end of the simulation, *i.e.* the time  $T_{max}$  corresponding to the



**Fig 6** Reference objects in the Utexture dataset.

biggest feature we want to detect, decreases, too, due to the larger time sampling step at the higher levels of the pyramid.

To locate extrema in the pyramid, we keep auxiliary images at each octave as done by DoG.<sup>9</sup> In particular, if we analyze  $L$  images at each octave, we actually compute  $L + 2$  images. We start an octave with the down-sampled version of the  $L + 1$ -th image of the previous octave and then let the wave simulation run for  $L + 2$  time steps. We then detect extrema by performing  $3 \times 3 \times 3$  non-maxima suppression on the  $L$  images from image 2 to image  $L + 1$ .

To perform the sharpness check on extrema, we create a 2D array of cumulative sums with the  $L$  logical images at each octave of the pyramid, *i.e.* by skipping the two auxiliary images at each octave. If  $t_1$  belongs to a different octave than  $t_2$ , we down-sample the cumulative sums at  $t_1$  to the size of the image at time  $t_2$  and we then use their difference to efficiently compute (24) and (23).

Finally, to recover the exact position in the original image of a keypoint detected at the higher octaves of the pyramid, we interpolate the wave evolution in space and time according to the method proposed by Brown and Lowe,<sup>43</sup> so to obtain the position and scale of the extrema with sub-pixel accuracy. Locating interest points to sub-pixel accuracy is especially important at higher levels of the pyramid, where the sampling step corresponds to larger offsets in the original image.

## 5 Experimental results

The proposed algorithm has been tested on three publicly available datasets. The first is the standard Oxford benchmark for detectors evaluation introduced in the well-know work by Mikolajczyk *et al.*<sup>15</sup> It includes 8 planar scenes and 5 nuisance factors: scale and rotation changes, viewpoint changes, decreasing illumination, blur and JPEG compression. Performance is measured according to repeatability and number of correct correspondences: repeatability between two images is computed as the ratio between the number of correct correspondences and the smaller number of detected regions located in the part of the scene present in both images; two regions form a correct correspondence if the intersection over union between them, when transformed according to the ground-truth homography, is greater then 50%.

The second dataset is the DTU Robot dataset.<sup>16</sup> It contains 60 scenes of planar and non-planar objects from a variety of categories such as miniature buildings, fabrics, and groceries. Each scene is captured from a set of repeatable positions along four paths by using a robotic arm. The dataset is somewhat complementary to the Oxford dataset in that it considers less kinds of nuisances (namely, scale changes, viewpoint changes and relighting), but comprises a more varied pool of scenes, therefore allowing to gain more insights on the general applicability of tested detectors. The authors consider only the recall rate, a performance figure analogous to the repeatability of the Oxford dataset.

Finally, to test the performance of detectors on untextured objects, we used the dataset introduced in,<sup>14</sup> which consists of 5 man-made objects (Fig. 6). The dataset focuses on scale and rotation invariance, but having been acquired with a hand-held camera, it also includes small out-of-plane rotations as well as slight defocus and light changes due to automatic parameter adjust-

**Table 2** Parameters for the linear models used to estimate  $t_{12}(r_{ext}) = a_t r_{ext} + b_t$  and  $\theta_{max}(r_{ext}) = a_\theta r_{ext} + b_\theta$  described in Fig. 5 and Sec. 3.5.

Method	$a_t$	$b_t$	$a_\theta$	$b_\theta$
Full Res INT 1/4 S	0.122	5.80	2.510	321.51
Full Res INT 1/6 S	0.143	6.40	2.401	335.21
Full Res MFI S	0.126	5.92	2.485	328.40
Full Res INT 1/4 NS	0.224	11.26	2.465	356.23
Full Res INT 1/6 NS	0.147	11.89	1.805	408.76
Full Res MFI NS	0.226	11.93	2.494	358.92
Pyramid INT 1/6 NS	0.410	6.231	0.113	380.82

ments performed by the camera driver.

We compare our proposal to a set of state-of-the-art detectors, whose code is publicly available. In particular, for DoG,<sup>9</sup> Harris-Affine, Harris-Laplace, Hessian-Affine, Hessian-Laplace,<sup>15</sup> IBR, EBR<sup>44</sup> and MSER<sup>23</sup> we used the binaries provided by the authors of;<sup>15</sup> for FastHessian<sup>21</sup> we used the original SURF code<sup>2</sup>.

As for parameters, we set  $k = 0.16$  and  $\rho = 0.07$ , *i.e.* we accept extrema at least as sharp as 7% of the maximum sharpness  $\theta_{max}$ , to allow low contrast symmetries to be detected in dark or overexposed images. We have experimentally found  $L = 16$  levels per octave to be a good trade-off between speed-up and simulation quality. We detect features as small as  $r_{min} = 6$  pixels (*i.e.* we start to search for extrema after  $n_{min} = \lfloor \frac{r_{min}}{\lambda} \rfloor$  iterations) and as big as  $r_{max} = 150$  pixels, (*i.e.* we run the simulation for  $n_{max} = \lceil \frac{r_{max}}{\lambda} \rceil$  iterations). We refer to our proposal as to WaDe (WAVE-based DETector).

### 5.1 Numerical scheme selection

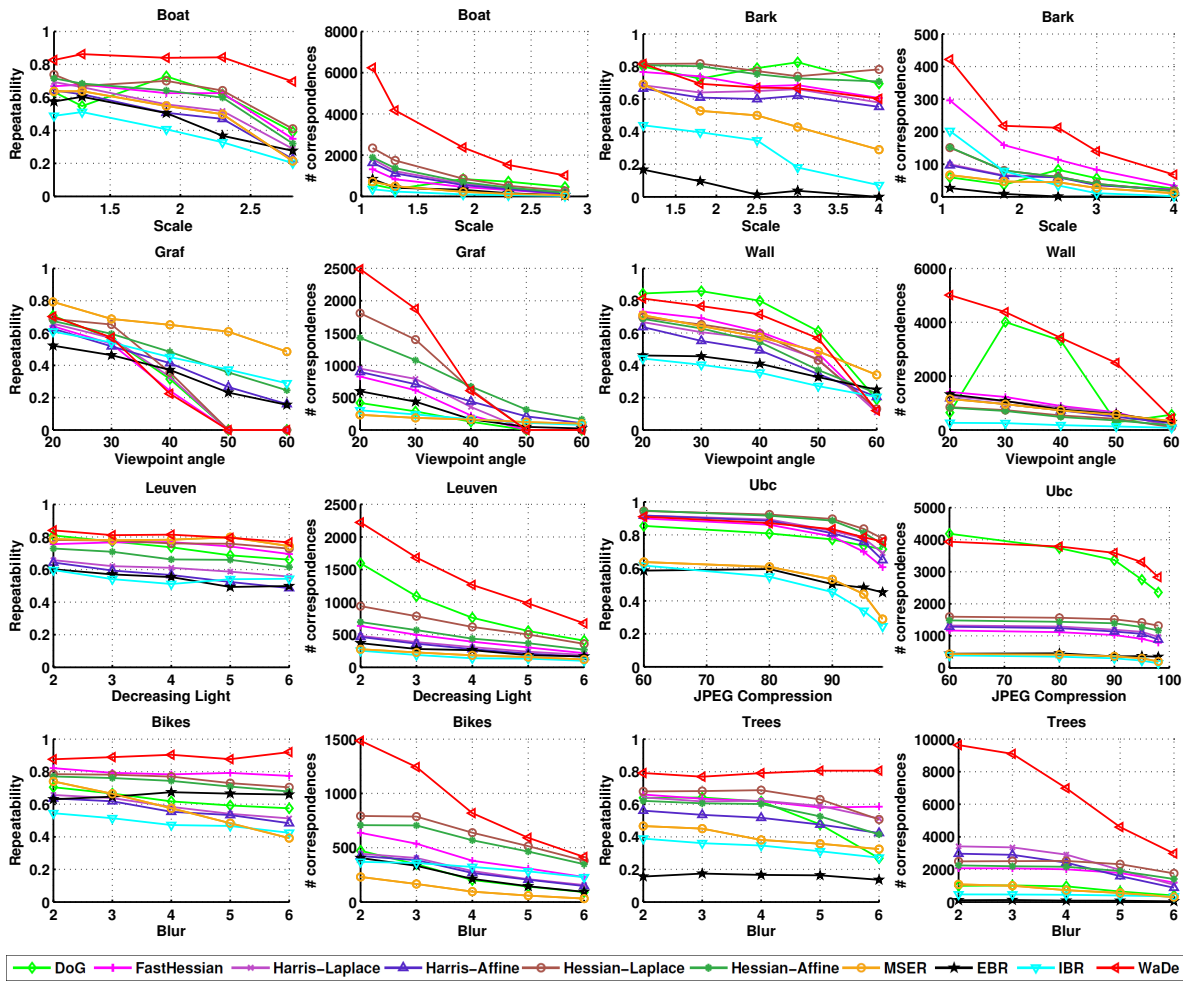
Before comparing our proposal to the other detectors, we carried out an experimental investigation concerning the most appropriate numerical scheme to simulate wave propagation within our algorithm. We performed the analysis without deploying the pyramidal approximation.

<sup>2</sup><http://www.vision.ee.ethz.ch/~surf/>



**Table 3** Comparison in terms of mean *repeatability* among the numerical schemes for wave simulation under different nuisances. In bold the best performance along each row, in italic the second best.

Nuisance	INT 1/4 NS	INT 1/4 S	INT 1/6 NS	INT 1/6 S	MFI NS	MFI S
Scale and Rotation	<b>0.89</b>	<i>0.88</i>	<b>0.89</b>	<i>0.88</i>	<b>0.89</b>	<i>0.88</i>
Decreasing Light	<i>0.82</i>	<i>0.82</i>	<b>0.83</b>	<i>0.82</i>	<i>0.82</i>	<i>0.82</i>
Viewpoint angle	<b>0.49</b>	<b>0.49</b>	<i>0.48</i>	<i>0.48</i>	<b>0.49</b>	<b>0.49</b>
JPEG Compression	<b>0.85</b>	<i>0.84</i>	<b>0.85</b>	<b>0.85</b>	<b>0.85</b>	<i>0.84</i>
Blur	<b>0.89</b>	<i>0.88</i>	<i>0.88</i>	<b>0.89</b>	<i>0.88</i>	<i>0.87</i>
Untextured	<i>0.35</i>	<i>0.36</i>	<b>0.37</b>	<i>0.36</i>	<i>0.35</i>	<i>0.36</i>



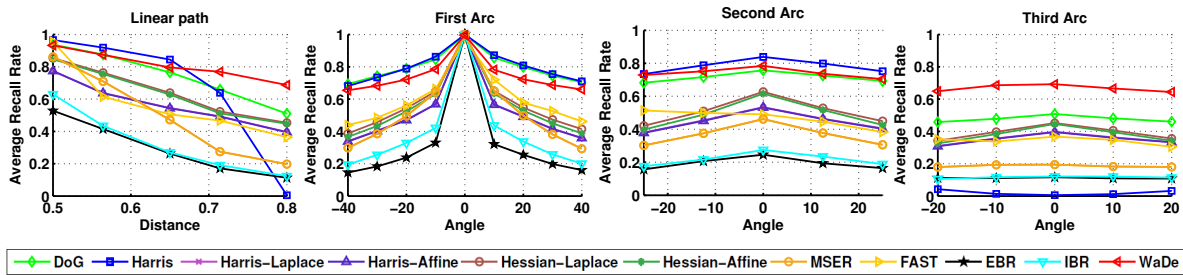
**Fig 7** Results on the Oxford dataset<sup>15</sup> for relative and absolute repeatability.

The neighborhood size and threshold of the sharpness test were estimated for all the considered schemes, both at the stability bound and at the middle of their stability range. The values of the regressed parameters are reported in Tab. 2. We then run WaDe using the different numerical

schemes on the Oxford and Untextured datasets, whose sizes allows to perform the comparison in a reasonable time. We report the mean repeatability for the different nuisances addressed by the datasets in Tab. 3. As already discussed with the help of Fig. 2, all the schemes are more isotropic and therefore more rotation invariant if used in the middle of their stability range (not-on-stability, NS in the table). Among the three NS schemes, INT 1/6 was shown to be less severely affected by numerical dispersion, *e.g.* the extreme diagonal spatial frequencies travel closer to the theoretical velocity than the corresponding frequencies in INT 1/4 and MFI schemes (see again Fig. 2). This is confirmed by the mean repeatability measurements reported in Tab. 3, as INT 1/6 NS turns out to provide overall the best performance across the nuisances addressed by the considered datasets. Surprisingly, an implicit scheme such as MFI, although far more complex to implement and expensive to run, does not deliver better performance. Based on these findings, we select INT 1/6 NS as the reference numerical scheme to simulate wave propagation in our keypoint detector and deploy its pyramidal approximation in all subsequent experiments using the regressed parameters reported in the last row of Tab. 2.

## 5.2 Oxford Dataset

Fig. 7 reports results on the full Oxford dataset. On the Boat and Bark sets of images, which test scale and rotation invariance, such theoretical properties of our algorithm are evidently confirmed. Moreover, the intuition that symmetries are a powerful cue to find repeatable keypoints is supported by the large margin in performance between our proposal and previous detectors on the Boat images. Leuven images allow for assessing robustness to lighting changes, and our method turns out slightly better than MSER, which is known to be the best detector in withstanding this nuisance.<sup>15</sup> On Graffiti and Wall images, which deal with robustness to affine deformations, our



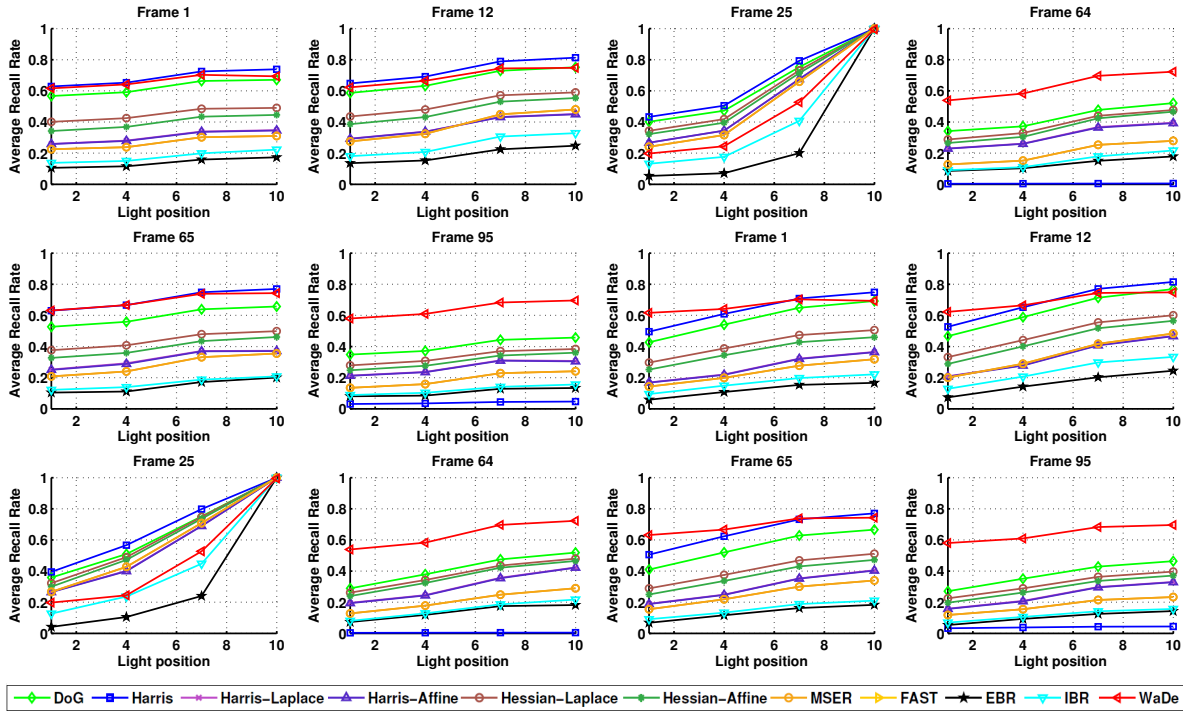
**Fig 8** Results on the Robot dataset<sup>16</sup> for camera movements along the linear path and the three arcs with diffuse lighting.

detector provides performance similar to DoG: one major drawback of both detectors is the lack of an affine renormalization of the extracted patch, which hinders their results in such comparison. In UBC images, the nuisance is JPEG compression: our detector turns out to be on par with the best previous detectors, namely Hessian-Laplace and Harris-Laplace, especially at higher compression rates. Finally, Bikes and Trees images show how symmetries are definitely the best cue to deal with blur: intuitively, symmetries are largely unaffected by isotropic smoothing.

As for the number of correspondences, our detector usually finds a higher number of repeatable keypoints than other methods. On some scenes the number is similar to DoG, which is usually slightly larger than the other methods, for instance on the UBC, Leuven, and Wall scenes. The order of magnitude is always similar to that of the other detectors, but on the Trees images: all the symmetries created by small leafs and details on the trunks are robustly captured by WaDe and results in an much higher number of repeatable keypoints.

### 5.3 Robot Dataset

Fig. 8 shows the first set of results on the Robot Dataset. In this experiment, we consider stable, diffuse illumination and camera movements along the linear path, which tests scale invariance, and the three arcs, which test robustness to a combination of scale and affine changes (the experiments



**Fig 9** Results on the Robot dataset<sup>16</sup> for changes in lighting with directional light. Every chart refer to a frame, which in turn implies a specific position of the robotic arm. In every chart, the horizontal axis reports the light position, from left to right in the first six charts, and from back to front in the remaining charts.

depicted in Fig. 9 in<sup>16</sup>). For memory and time constraints (the whole dataset amounts to 730 GB), we run our proposal only on the reduced dataset selected by the authors. Hence results are available only at a reduced number of sample points. Results for all other detectors were kindly provided by the authors of.<sup>16</sup> As far as the linear path is concerned, the proposed algorithm shows the best robustness to scale variation and outperforms by a large margin all the other detectors at the maximum scale variation. This result confirms the findings on the Oxford Boat and Bark dataset and reinforces them by demonstrating that they hold for a variety of scenes. As for the arcs, on the first two arcs our detector is on par with DoG and Harris, the best performers, and outperforms all the tested detectors on the third arc.

Fig. 9 reports the second experiment on the Robot Dataset, dealing with lighting changes. In

these experiments, a directional light is simulated according to the protocol proposed in,<sup>16</sup> *i.e.* by weighing the images acquired with a single led turned on, so as to simulate a directional light moving either from left to right or from back to front around the scene (the experiments depicted in Fig. 10 and 11 in<sup>16</sup>). WaDe provides overall the best performance, being on par with Harris when scale changes are limited, and outperforming all the tested detectors on frames 64 and 95, *i.e.* when the camera is moved to the third, furthest arc.

#### 5.4 *Untextured Dataset*

Fig. 10 depicts results on the Untextured objects dataset. We were unable to run IBR and EBR on all set of images as the binaries crashed on some sets: yet, performance on the remaining sets was significantly worse than others, so we do not report them in the charts. Untextured objects set forth significant difficulties for current state-of-the-art detectors: the number of correct correspondences drops by two orders of magnitude with respect to the other datasets, and repeatability falls below 50%. Please note that, as done in,<sup>14</sup> on this dataset we lower from 40% to 20% the overlap error to consider two keypoints as repeatable, as we found that this is the maximum error that a robust descriptor as SIFT can tolerate with poorly textured objects. Symmetries confirm to be a robust cue even in this challenging scenario: their repeatability is the least affected by the increasing difficulties of the scenes. They also turn out to provide the highest number of correspondences, an important practical trait to enable their use in recognition of untextured objects in the presence of occlusions.

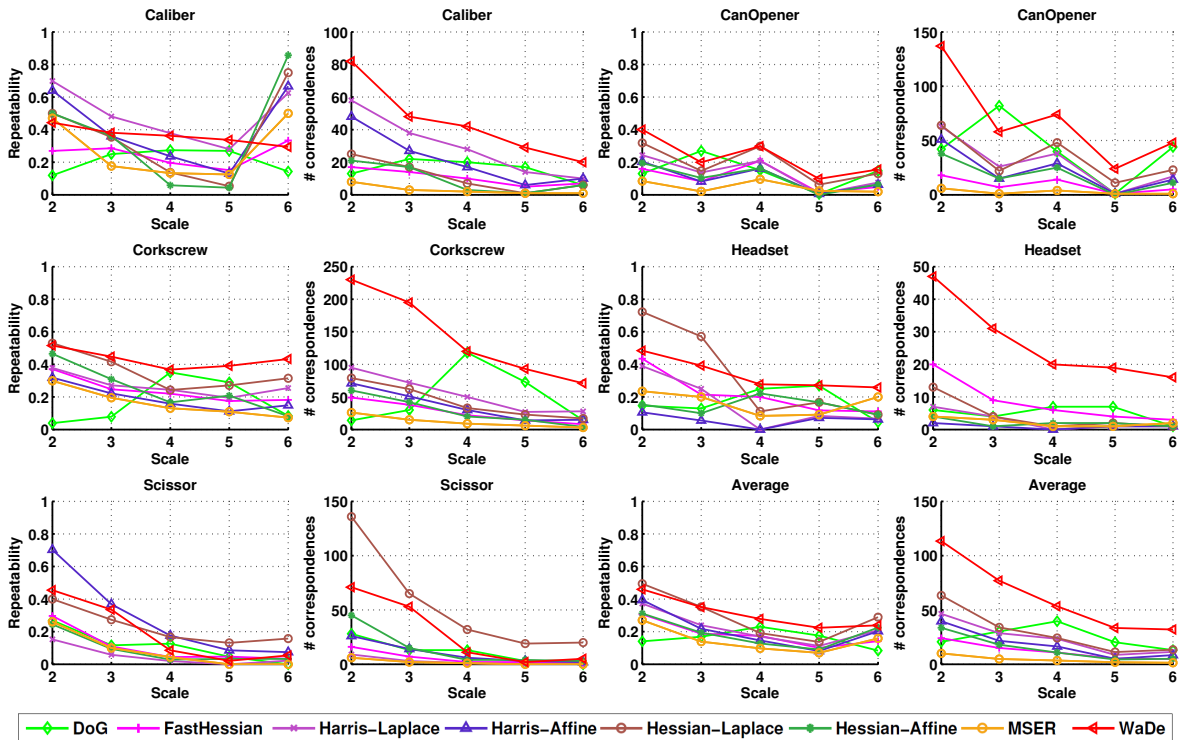


Fig 10 Results on the Untextured dataset<sup>14</sup> for relative and absolute repeatability.

### 5.5 Descriptor matching

Using again the Oxford dataset, we assess the distinctiveness of the detected repeatable regions by considering how well can they be matched by a state-of-the-art descriptor such as SIFT.<sup>9</sup> Although favouring detectors either proposed together with the selected descriptor or designed to take advantage of its traits, this methodology is regarded as the standard approach<sup>15</sup> to ascertain the quality of the regions found by a detector due to the widespread use of SIFT.

Results are reported in Fig. 11. The performance of the descriptor are greatly influenced by both the content of images and the transformation under test in each set, the overall performance varying between consistently low matching scores, as in Bark and Graffiti, and always very good results, as in UBC. In spite of this variability, the overall ranking of the descriptiveness of the regions found by the tested detectors is consistent. Regions by WaDe and DoG turn out to be

the most distinctive. In particular, they are: as good as the others when the descriptor is not particularly effective, as in Bark or Graffiti; on par on some scenes (Leuven, Trees); the best for a particular scene, as it is the case of DoG for Boat or UBC and WaDe for Bikes or Wall. Overall, WaDe regions can be used to obtain point-to-point correspondences by matching SIFT descriptors with a performance level comparable to DoG, which was proposed together with SIFT. This fact, combined with the observation that the number of correct matches is always equivalent or higher for WaDe with respect to the other detectors, vouches that WaDe keypoints offer not only state-of-the-art repeatability but also remarkable distinctiveness, and, as such, are amenable to effective deployment within practical local feature pipelines.

### *5.6 Comparison to DoG*

As suggested by a reviewer, we present here a brief comparative discussion between WaDe and DoG motivated by prominence of the latter method in the local invariant features literature. By comparing the red lines (WaDe) to the green lines (DoG) in Figures 7-8-9, we can notice how WaDe is significantly more robust to blur (Fig. 7, last row) as well as, on average, to scale changes (Fig. 7. Boat and Bark, Fig.9 linear path). Besides, WaDe seems to provide overall better performance in varying lighting conditions (Fig. 7, Leuven and Fig. 9). On the other hand, WaDe and DoG behave similarly in terms of robustness to out-of-plane rotations (Fig. 7 - Graf and Wall, Fig. 8 first, second and third arc). When it comes to untextured scenes (Fig. 10), WaDe outperforms DoG in all experiments, which supports our intuition on symmetries providing more robust and distinctive cues in these settings. It is also worth pointing out how WaDe yields quite consistently a notably larger number of repeatable keypoints than DoG (Figs. 7 and 10, charts depicting #

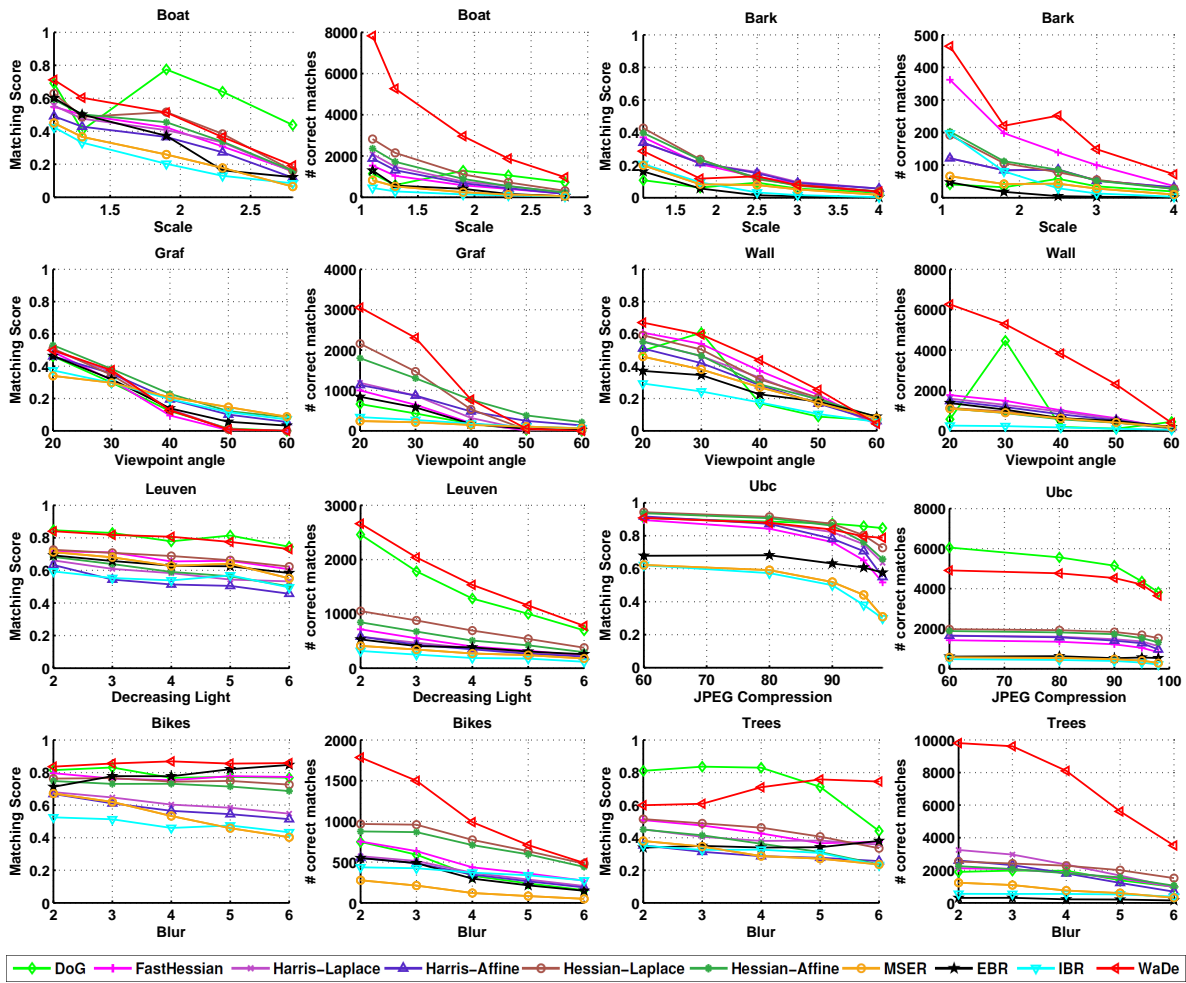


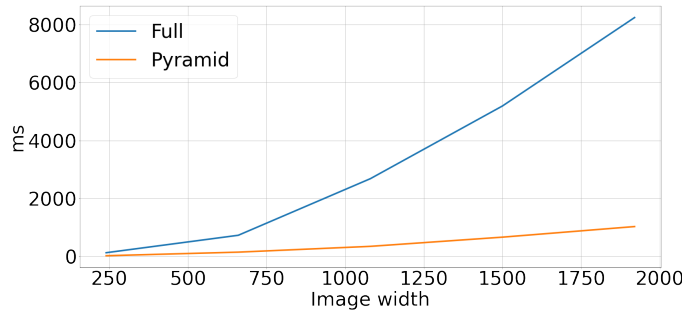
Fig 11 Results on the Oxford dataset<sup>15</sup> for descriptor matching experiments.

correspondences), a very useful property in practical settings. Eventually, we found it remarkable that WaDe can detect keypoints as effective as DoG to establish correspondences by matching SIFT descriptors (Fig. 11) despite DoG and SIFT having been proposed as two tightly coupled steps of the same local invariant features pipeline.

### 5.7 Speed-up by pyramidal approximation

Finally, we evaluate the gain in execution time obtained by introducing the pyramidal approximation. Indeed, it is noteworthy that all previous results have been obtained with such an approxima-





**Fig 12** Execution times of the pyramidal approximation and the full resolution algorithm as a function of image size.

tion in place; they already show that its deployment does not reduce the performance with respect to the full resolution algorithm. Yet, the pyramid has a positive impact on the efficiency of the algorithm. As shown in Fig. 12, while the full resolution algorithm run-time grows exponentially with the image size, the pyramid approximation increase is almost linear with the image size, and for large images can be faster than the full resolution algorithm by almost an order of magnitude. The C++ implementation of the pyramidal approximation can process a  $640 \times 480$  images (Boat from the Oxford dataset) at 7 frames per second on a standard PC without any optimization, whereas the full resolution algorithm runs at 1 frame per second on the same image.

## 6 Concluding remarks

This work shows that the traditional linear scale-space theory is not the unique way to address scale-invariant feature detection within an elegant conceptual framework. Instead of the heat equation, hyperbolic PDEs, and in particular the wave equation, can indeed be used to derive a one-parameter family of signals that enables scale-invariant image analysis. The wave equation fires on symmetries, and we have shown they can provide robust and discriminative cues for detection of repeatable keypoints on a broader set of image structures than those found on highly textured surfaces. We have discussed the problem of numerical simulation of wave propagation in images

and selected the most appropriate scheme and simulation parameters for our peculiar use. We have also shown how to speed-up the algorithm without penalizing its performance by deploying a pyramid of images where the Courant number is kept constant between octaves.

As suggested by prominent evaluations,<sup>15,16</sup> salient regions are usually complementary, so that a set of diverse detectors might be deployed whenever one cannot predict in advance the likely scene structures. The proposed algorithm, based on a rather overlooked cue such as symmetry, might turn out highly complementary with respect to current state-of-the-art detectors. As vouched by earlier work,<sup>11</sup> our approach is nicely implementable on fine grained data-parallel architectures, and as such holds the potential for effective acceleration via GPU processing.

### 6.1 Future work

Despite good results, we feel that there is still room for improvements. Different methods or domains for the simulation of waves having better numerical properties may be considered and their impact on the overall performance and run-time of the algorithm be investigated, *e.g.* more robust but less efficient numerical schemes<sup>45</sup> or hexagonal lattices.<sup>46</sup> Moreover, several complementary studies may be envisaged, in particular, in order to either design or learn a tightly coupled descriptor computed directly from wave simulation. Another possible avenue for future work deals with extending the detection of wave-based features to color images, *e.g.* via the model of images as manifold embedded in a combined spatial-color space.<sup>47</sup>

### References

- 1 P. Perona and J. Malik, “Scale-space and edge detection using anisotropic diffusion,” *IEEE Trans. Pattern Anal. Mach. Intell.* **12**, 629–639 (1990).

- 2 M. Lysaker, A. Lundervold, and X. Tai, “Noise removal using fourth-order partial differential equations with applications to medical magnetic resonance images in space and time,” *IEEE Trans. Img. Proc.* **12**, 1579–1590 (2003).
- 3 R. H. Chan, A. Lanza, S. Morigi, *et al.*, “An adaptive strategy for the restoration of textured images using fractional order regularization,” *Numerical Mathematics: Theory, Methods and Applications* **6**, 276–296 (2013).
- 4 S. Osher and L. I. Rudin, “Feature-oriented image enhancement using shock filters,” *SIAM J. Numer. Anal.* **27**, 919–940 (1990).
- 5 V. Ratner and Y. Y. Zeevi, “Denoising-enhancing images on elastic manifolds,” *Trans. Img. Proc.* **20**, 2099–2109 (2011).
- 6 A. P. Witkin, “Scale-space filtering,” in *Proc. Eighth Int. Joint Conf. on Artificial Intelligence, IJCAI’83* **2**, 1019–1022 (1983).
- 7 J. Koenderink, “The structure of images,” *Biological Cybernetics* **50**(5), 363–370 (1984).
- 8 T. Lindeberg, *Scale-Space Theory in Computer Vision*, Kluwer Academic Publishers, Norwell, MA, USA (1994).
- 9 D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vision* **60**, 91–110 (2004).
- 10 K. Hanahara and M. Hiyane, “A circle-detection algorithm simulating wave propagation,” *Mach. Vision Appl.* **4**, 97–111 (1991).
- 11 R. Cucchiara, L. Di Stefano, and M. Piccardi, “Detection of circular objects by wave propagation on a mesh-connected computer,” *Journal of Parallel and Distributed Computing* **31**(1), 77–87 (1995).

- 12 G. L. Scott, S. C. Turner, and A. Zisserman, "Using a mixed wave/diffusion process to elicit the symmetry set," *Image Vision Comput.* **7**, 63–70 (1989).
- 13 H. Tek and B. B. Kimia, "Symmetry maps of free-form curve segments via wave propagation," *Int. J. Comput. Vision* **54**, 35–81 (2003).
- 14 S. Salti, A. Lanza, and L. D. Stefano, "Keypoints from symmetries by wave propagation," in *Comp. Vision and Pattern Recog. (CVPR), 2013 IEEE Conference on*, 2898–2905 (2013).
- 15 K. Mikolajczyk, T. Tuytelaars, C. Schmid, *et al.*, "A comparison of affine region detectors," *Int. J. Comput. Vision* **65**, 43–72 (2005).
- 16 H. Aanæs, A. L. Dahl, and K. Steenstrup Pedersen, "Interesting interest points," *Int. J. Comput. Vision* **97**, 18–35 (2012).
- 17 T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: a survey," *Found. Trends. Comput. Graph. Vis.* **3**, 177–280 (2008).
- 18 Y. Li, S. Wang, Q. Tian, *et al.*, "A survey of recent advances in visual feature detection," *Neurocomputing* (0), 736 – 751 (2015).
- 19 H. P. Morevec, "Towards automatic visual obstacle avoidance," in *Proceedings of the 5th International Joint Conference on Artificial Intelligence - Volume 2, IJCAI'77*, 584–584, Morgan Kaufmann Publishers Inc., (San Francisco, CA, USA) (1977).
- 20 C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. Fourth Alvey Vision Conf.*, 147–151 (1988).
- 21 H. Bay, A. Ess, T. Tuytelaars, *et al.*, "Speeded-up robust features (SURF)," *Comput. Vis. Image Underst.* **110**, 346–359 (2008).

- 22 K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *Int. J. Comput. Vision* **60**(1), 63–86 (2004).
- 23 J. Matas, O. Chum, M. Urban, *et al.*, "Robust wide baseline stereo from maximally stable extremal regions," in *Proc. British Machine Vision Conference, BMVC'02* **1**, 384–393 (2002).
- 24 E. Rosten and T. Drummond, "Fusing points and lines for high performance tracking," in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, **2**, 1508–1515 Vol. 2 (2005).
- 25 E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *In European Conference on Computer Vision*, 430–443 (2006).
- 26 E. Rosten, R. Porter, and T. Drummond, "Faster and better: a machine learning approach to corner detection," *IEEE Trans. Pattern Anal. Mach. Intell.* **32**, 105–119 (2010).
- 27 E. Mair, G. D. Hager, D. Burschka, *et al.*, "Adaptive and generic corner detection based on the accelerated segment test," in *Proceedings of the European Conference on Computer Vision (ECCV'10)*, (2010).
- 28 S. Leutenegger, M. Chli, and R. Siegwart, "Brisk: Binary robust invariant scalable keypoints," in *Computer Vision (ICCV), 2011 IEEE International Conference on*, 2548–2555 (2011).
- 29 E. Rublee, V. Rabaud, K. Konolige, *et al.*, "Orb: An efficient alternative to sift or surf," *Computer Vision, IEEE International Conference on* **0**, 2564–2571 (2011).
- 30 W. Förstner, T. Dickscheid, and F. Schindler, "Detecting interpretable and accurate scale-invariant keypoints," in *Computer Vision, 2009 IEEE 12th International Conference on*, 2256–2263 (2009).

- 31 P. Mainali, G. Lafruit, Q. Yang, *et al.*, “Sifer: Scale-invariant feature detector with error resilience,” *Int. J. Comput. Vision* **104**(2), 172–197 (2013).
- 32 P. F. Alcantarilla, A. Bartoli, and A. J. Davison, “KAZE features,” in *Eur. Conf. on Computer Vision (ECCV)*, (2012).
- 33 P. F. Alcantarilla, J. Nuevo, and A. Bartoli, “Fast explicit diffusion for accelerated features in nonlinear scale spaces,” in *British Machine Vision Conf. (BMVC)*, (2013).
- 34 J. Maver, “Self-similarity and points of interest,” *IEEE Trans. Pattern Anal. Mach. Intell.* **32**, 1211–1226 (2010).
- 35 D. Hauagge and N. Snavely, “Image matching using local symmetry features,” in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 206–213 (2012).
- 36 D. Reisfeld, H. Wolfson, and Y. Yeshurun, “Context-free attentional operators: the generalized symmetry transform,” *Int. J. Comput. Vision* **14**, 119–130 (1995).
- 37 G. Loy and A. Zelinsky, “Fast radial symmetry for detecting points of interest,” *IEEE Trans. Pattern Anal. Mach. Intell.* **25**, 959–973 (2003).
- 38 B. Engquist and A. Majda, “Absorbing boundary conditions for the numerical simulation of waves,” *Mathematics of Computation* **31**, 629–651 (1977).
- 39 A. R. Gourlay and A. R. Mitchell, “A classification of split difference methods for hyperbolic equations in several space dimensions,” *SIAM J. Numer. Anal.* **6**, 62–71 (1969).
- 40 J. Douglas, Jr., “On the numerical integration of  $u_{xx} + u_{yy} = u_t$  by implicit methods,” *Journal of the Society of Industrial and Applied Mathematics* **3**, 42–65 (1955).
- 41 M. van Walstijn and K. Kowalczyk, “On the numerical solution of the 2d wave equation with compact fdtd schemes,” in *Proc. Digital Audio Effects (DAFx)*, 205–212 (2008).

- 42 M. Van Walstijn and K. Kowalczyk, “On the numerical solution of the 2D wave equation with compact FDTD schemes,” in *Proc. Int. Conf. Digital Audio Effects (DAFX08)*, 205–212 (2008).
- 43 M. Brown and D. Lowe, “Invariant features from interest point groups,” in *In British Machine Vision Conference*, 656–665 (2002).
- 44 T. Tuytelaars and L. Van Gool, “Matching widely separated views based on affine invariant regions,” *Int. J. Comput. Vision* **59**, 61–85 (2004).
- 45 D. Yang, J. Peng, M. Lu, *et al.*, “Optimal nearly analytic discrete approximation to the scalar wave equation,” *Bulletin of the Seismological Society of America* **96**(3), 1114–1130 (2006).
- 46 B. Hamilton and S. Bilbao, “Hexagonal vs. rectilinear grids for explicit finite difference schemes for the two-dimensional wave equation,” *The Journal of the Acoustical Society of America* **133**(5), 3532–3532 (2013).
- 47 N. Sochen and Y. Y. Zeevi, “Images as manifolds embedded in spatial-feature non-euclidean space,” Tech. Rep. CCIT260, Israel Inst. Technol, Haifa, Haifa, Israel (1998).

**Samuele Salti** received the M.Sc. degree in computer engineering in 2007 and the Ph.D. degree in computer engineering in 2011, both from the University of Bologna, Italy. Since 2018 he is assistant professor at the Department of Computer Science and Engineering of the University of Bologna. In 2007 he visited the Heinrich-Hertz-Institute in Berlin, Germany and in 2010 the Multimedia and Vision Research Group (MMV) at Queen Mary, University of London. His research interests are deep learning applied to computer vision applications, especially 3D computer vision. Dr. Salti has co-authored 40 publications in international conferences and journals. He was awarded the best paper award runner-up at 3DIMPVT in 2011. He is a member of CVPL.