

Keypoints from Symmetries by Wave Propagation

Samuele Salti
University of Bologna

samuele.salti@unibo.it

Alessandro Lanza
University of Bologna

alessandro.lanza2@unibo.it

Luigi Di Stefano
University of Bologna

luigi.distefano@unibo.it

Abstract

The paper conjectures and demonstrates that repeatable keypoints based on salient symmetries at different scales can be detected by a novel analysis grounded on the wave equation rather than the heat equation underlying traditional Gaussian scale-space theory. While the image structures found by most state-of-the-art detectors, such as blobs and corners, occur typically on planar highly textured surfaces, salient symmetries are widespread in diverse kinds of images, including those related to untextured objects, which are hardly dealt with by current feature-based recognition pipelines. We provide experimental results on standard datasets and also contribute with a new dataset focused on untextured objects. Based on the positive experimental results, we hope to foster further research on the promising topic of scale invariant analysis through the wave equation.

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. Purposely, 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 [16], fourth-order [12] and, very recently, fractional-order PDEs [3] 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 the shock-filters [15] or the telegrapher equation [17].

On the other hand, 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. The scale-space theory was developed in seminal works by Witkin [24] and Koenderink

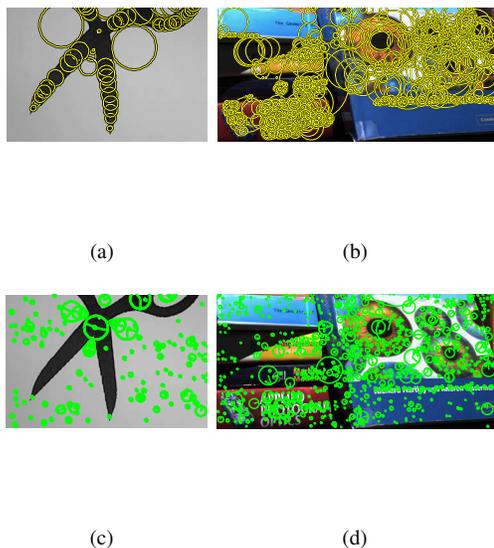


Figure 1: (a-b) keypoints by our algorithm; (c-d) keypoints by DoG [10]. Symmetries are more likely than blobs to appear in untextured objects (a-c): only a few DoG keypoints actually lay on the object. On the other hand, a similar number of keypoints is detected on textured surfaces (b-d): yet, the proposed keypoints concentrate on more evident (*i.e.* likely more repeatable) structures.

[8], and later popularized by the automatic scale selection principle proposed by Lindeberg [9]. Lindeberg's scale-normalized derivatives provide the theoretical foundations for the Differences-of-Gaussians (DoG) keypoint detector [10], arguably among the most influential contributions in the field of computer vision. Despite such a relevant result, 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 [6, 4, 19], 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 [21].

In this paper we investigate on the use of PDEs in image analysis from the modern perspective of local invariant features. In particular, we focus on the problem of 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 (Fig. 1), 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 (Fig. 1). Thus, the proposed approach qualifies as a general tool for repeatable salient regions detection. Surprisingly, there has been relatively little work on the use of symmetry as a cue to detect and describe local features. The most recent contribution on this topic is due to Haugge et Snavely [7], who propose 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 [18, 11] 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.

2. The wave equation

The wave equation is a well-known linear second-order hyperbolic partial differential equation. It is used in several important 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:

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

where $c \in \mathbb{R}^+$ is the wave speed, ∇^2 is the 2-D spatial Laplacian operator, 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. 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 [5], which are inherently non-local in space and time, *i.e.* to advance one time level at a single point they require information over 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 the problem (1) well-posed [5].

To numerically solve (1) we use a fully explicit finite difference scheme based on the following standard approximations of second order time and space partial derivatives:

$$\begin{aligned} u_{tt}(x_i, y_j, t^n) &\simeq (u_{i,j}^{n+1} - 2u_{i,j}^n + u_{i,j}^{n-1}) / \Delta t^2 & (2) \\ \nabla^2 u(x_i, y_j, t^n) &\simeq (u_{i-1,j-1}^n + 2u_{i-1,j}^n + u_{i-1,j+1}^n + & (3) \\ &2u_{i,j-1}^n - 12u_{i,j}^n + 2u_{i,j+1}^n + \\ &u_{i+1,j-1}^n + 2u_{i+1,j}^n + u_{i+1,j+1}^n) / 4\Delta x^2 \end{aligned}$$

on the regular space-time domain mesh $y_i = i \Delta y$, $x_j = j \Delta x$, $t^n = n \Delta t$. Letting the spatial mesh coincide with the pixel grid ($\Delta x = \Delta y = 1$, $i = 1, \dots, h$, $j = 1, \dots, w$, with w and h the image width and height in pixels, respectively), denoting by $r = c\Delta t/\Delta x \in \mathbb{R}^+$ the Courant number and substituting the above finite difference formulae into the wave equation, we obtain the following explicit time-marching scheme:

$$\begin{aligned} u_{i,j}^{n+1} &= \frac{r^2}{4} (u_{i-1,j-1}^n + 2u_{i-1,j}^n + u_{i-1,j+1}^n + & (4) \\ &2u_{i,j-1}^n - 12u_{i,j}^n + 2u_{i,j+1}^n + \\ &u_{i+1,j-1}^n + 2u_{i+1,j}^n + u_{i+1,j+1}^n) + 2u_{i,j}^n - u_{i,j}^{n-1} \end{aligned}$$

for 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 remaining nodes, we exploit the initial conditions, whose discretization easily yields an explicit solution for the inner nodes at the first iteration:

$$\begin{aligned} u_{i,j}^1 &= \frac{r^2}{8} (u_{i-1,j-1}^0 + 2u_{i-1,j}^0 + u_{i-1,j+1}^0 + & (5) \\ &2u_{i,j-1}^0 - 12u_{i,j}^0 + 2u_{i,j+1}^0 + \\ &u_{i+1,j-1}^0 + 2u_{i+1,j}^0 + u_{i+1,j+1}^0) + u_{i,j}^0 \end{aligned}$$

as well as the boundary conditions, whose discretization yields an explicit solution for the image boundary nodes:

$$\begin{aligned} u_{i,1}^{n+1} &= u_{i,1}^n + r(u_{i,2}^n - u_{i,1}^n) \\ u_{1,j}^{n+1} &= u_{1,j}^n + r(u_{2,j}^n - u_{1,j}^n) \\ u_{i,w}^{n+1} &= u_{i,w}^n + r(u_{i,w-1}^n - u_{i,w}^n) \\ u_{h,j}^{n+1} &= u_{h,j}^n + r(u_{h-1,j}^n - u_{h,j}^n). \end{aligned} \quad (6)$$

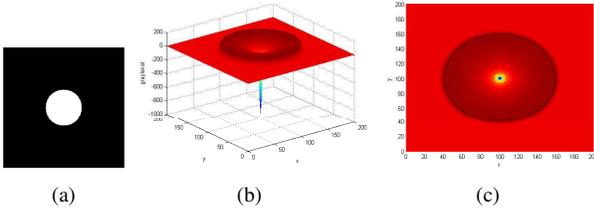


Figure 2: (a) initial image I ; (b-c) function $u(x, y; t)$, from a side and a top view respectively, at the time t when the wave has traveled a distance equal to the radius of the circle (graylevel intensities get higher from blue to red).

Numerical stability of the proposed explicit scheme constrains 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. To obtain a stable scheme as well as an easily interpretable family of output signals, we fix $c\Delta t = \frac{1}{2}$, *i.e.* the wave travels half pixel every Δt . In particular, we obtain this by setting $c = \frac{\sqrt{2}}{2}$, $\Delta t = \frac{\sqrt{2}}{2}$.

Despite this constraint on Δt , we can use an explicit scheme because we are not interested in computing with few iterations the solution at large times, but rather in analyzing the wave evolution with time steps commensurable to the image lattice. Such a scheme results in an efficient algorithm.

3. Wave-based scale analysis

The evolution of image intensities obtained through the wave equation (1) allows for multi-scale signal analysis. In particular, as shown in [6], 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. 2a, and think of it as a wave 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 shown by numerical simulation in Fig. 2b-2c.

Detection of circle centers is actually a particular case of the more general property of wave propagation of eliciting the symmetry set of curves [19]. 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 at each iteration the local spatial extrema of the accumulator into the set. Such approach detects all kind of symmetries, regardless of their saliency, and is effective to skeletonize binary shapes. To define repeatable keypoints, only the points of locally maximal symmetry must be considered, which correspond to the shock

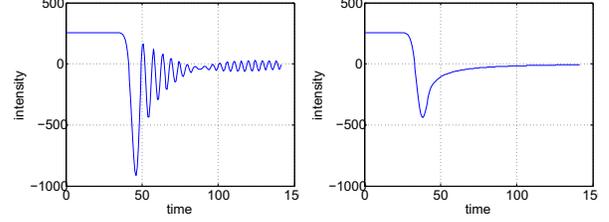


Figure 3: Temporal evolution of the intensity at the center of the circle in Fig. 2a under the discretized wave process (left) and the discretized wave-diffusion process (right).

points of the wave propagation (see [21] 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-invariant 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. Indeed, one interesting future work is the theoretical investigation on the definition of a sound scale-space theory from the wave equation.

Unfortunately, discretization of the wave propagation is prone to quite significant numerical errors, mainly due to numerical (or grid) dispersion [23]. This can be clearly seen in Fig. 3, left, where numerical dispersion causes spurious oscillations. To overcome such issues, we adopt a solution similar to that proposed in [19]: we interleave a wave propagation step to a linear diffusion step governed by the heat equation with diffusivity $k \in \mathbb{R}^+$,

$$u_t(x, y; t) = k\nabla^2 u(x, y; t), \quad (7)$$

which we discretize using, for the spatial Laplacian, the same scheme (3) used for the wave equation and, for the time derivative, forward differences. These choices result in the following explicit scheme, that we interleave with (4),

$$u_{i,j}^{n+1} = \frac{p}{4} (u_{i-1,j-1}^{n+\frac{1}{2}} + 2u_{i-1,j}^{n+\frac{1}{2}} + u_{i-1,j+1}^{n+\frac{1}{2}} + 2u_{i,j-1}^{n+\frac{1}{2}} - 12u_{i,j}^{n+\frac{1}{2}} + 2u_{i,j+1}^{n+\frac{1}{2}} + u_{i+1,j-1}^{n+\frac{1}{2}} + 2u_{i+1,j}^{n+\frac{1}{2}} + u_{i+1,j+1}^{n+\frac{1}{2}}) + u_{i,j}^{n+\frac{1}{2}} \quad (8)$$

where $p = \frac{k\Delta t}{\Delta x^2}$ and $u^{n+\frac{1}{2}}$ is the output of the last wave propagation step. We found experimentally that the previous choice of $c = \Delta t = \frac{\sqrt{2}}{2}$ together with $k = 0.16$ guarantees stability also of the interleaved process. The beneficial effect of the wave-diffusion process is clearly visible in Fig. 3, right, where spurious oscillations have been filtered out.

4. Extrema detection and validation

As can be seen in Fig. 4, the spatio-temporal extrema of $u(x, y; t)$ identify perceptually meaningful structures of

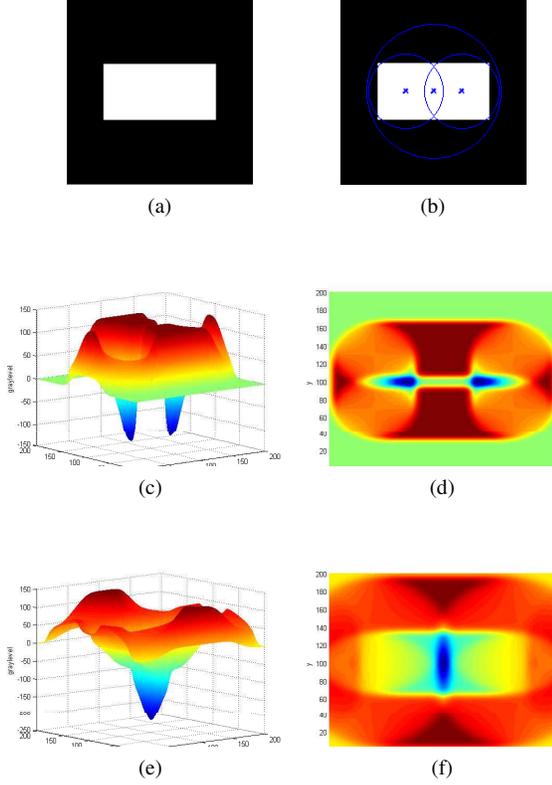


Figure 4: (a) initial image I ; (b) extrema of $u(x, y; t)$; (c-d) wave state when the two smaller keypoints are detected, from a side and a top view, respectively; (e-f) wave state when the third keypoint is detected.

images: 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 an 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:

$$r_{ext} = \frac{c t_{ext}}{\Delta x} = \frac{c \Delta t n_{ext}}{\Delta x}. \quad (9)$$

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

Given such scale covariant extrema in our family of signals, we define as keypoints the *sharp* local extrema of $u(x, y; t)$. Although on the synthetic images showed 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.

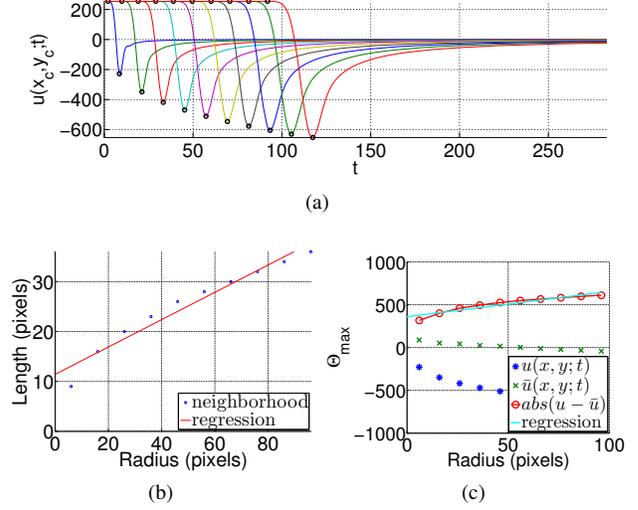


Figure 5: Definition of sharp extrema: (a) Temporal evolution of the intensity at the center (x_c, y_c) of circles of varying radius. The extrema (minima) show up at times proportional to the circles radius. Black dots on each curve indicate the neighborhood used at each scale for sharpness-based keypoints 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.

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

$$|u(x_{ext}, y_{ext}, t_{ext}) - \bar{u}(x_{ext}, y_{ext}, t_{ext})| \geq \theta, \quad (10)$$

$$\bar{u}(x_{ext}, y_{ext}, t_{ext}) = \frac{1}{t_2 - t_1} \sum_{k=t_1}^{t_2} u(x_{ext}, y_{ext}, k). \quad (11)$$

Both the neighborhood $[t_1, t_2]$ and the threshold θ must adapt with scale, as the absolute value at the extrema as well as their smoothness increase with scale (Fig. 5a). 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 edge of the image and the computation of the average is computationally more demanding 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 must be small enough to avoid the influence of clutter and occlusions in real applications, but also large enough to allow discriminating between weak



Figure 6: Reference objects in the untextured dataset.

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 t_1 of the neighborhood, we analyze the behavior of the most symmetrical shape with the highest contrast, *i.e.* a circle like that depicted in Fig. 2a. 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. Therefore, we define the starting point of the neighborhood as the end of the plateau (Fig. 5a). As the trend deviates more and more from the continuous ideal solution when the scale gets larger, because of the numerical dispersion and the diffusion that we interleave with propagation, we learn a starting point for each scale by regressing the neighborhood of the circle (Fig. 5b). In particular, a linear model seems appropriate and gives the following equation,

$$t_1(r_{ext}) = 0.274r_{ext} + 11.43. \quad (12)$$

Finally, using again the same image, we regress the value of the absolute difference $|u - \bar{u}|$ (Fig. 5c): this defines the maximum meaningful value θ_{max} for the threshold θ . We choose again a linear model, which yields

$$\theta_{max}(r_{ext}) = 2.95r_{ext} + 360. \quad (13)$$

As we are interested in detecting weaker, though repeatable, symmetries than a high-contrast circle, the actual value θ used at each scale is defined as a fraction ρ of $\theta_{max}(r_{ext})$.

The domain for searching local extrema must also take into account the peculiarity of the discretized wave propagation. The most straightforward estimation of local extrema of 3D continuous signals, given their discretized version, is represented by extrema within a $3 \times 3 \times 3$ neighborhood in the discrete domain. However, in wave propagation time and space discretizations are bound by the constant c , which means, in particular, that it takes $\frac{1}{c\Delta t}$ discrete time intervals for the wave to travel 1 pixel. Therefore, to correctly determine if the wave state at (x, y, t) is a local extremum we have to compare it not only with its 3×3 neighbors at time $t - 1$ and $t + 1$, but with all the 3×3 neighbors at time $k \in [t - \frac{1}{c\Delta t}, t + \frac{1}{c\Delta t}]$. In other words, to read the correct wave state in past and future times we have to give

to all the contributions the time to move to a given pixel from its neighbors, process that can take at most $\frac{1}{c\Delta t}$ time steps. With our choice of Δt and c , we have $\frac{1}{c\Delta t} = 2$, and, hence, it turns out that we have to perform the search for local extrema in a $3 \times 3 \times 5$ neighborhood. The overall detector is summarized in Alg. 1.

5. Experimental results

The proposed algorithm has been tested on two publicly available datasets. The first is the standard Oxford benchmark for detectors evaluation introduced in the well-know comparison by Mikolajczyk *et al.* [14]. 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.

The second dataset is the recently introduced DTU Robot dataset [1]. 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 nuisances (namely, only 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, an analogous of repeatability in the previous dataset.

Moreover, to test the performance of detectors on challenging, untextured objects, we introduce a novel dataset consisting 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 adjustments done by the camera driver. We estimated the ground-truth homographies between reference and test images and then used the evaluation tool provided together with the Oxford dataset.

An unoptimized C++ implementation of our detector runs at about 1.5 frames/second on 640×480 images. Such implementation, together with our untextured objects dataset are publicly available at the project website¹. We compare our proposal to a set of state-of-the-art detectors, whose code is publicly available. In particular, for DoG [10], Harris-Affine, Harris-Laplace, Hessian-Affine, Hessian-Laplace [14], IBR, EBR [22] and MSER [13] we used the binaries provided by the authors of [14]; for FastHessian [2] we used the original SURF code².

As for parameters, we use $\Delta t = c = \frac{\sqrt{2}}{2}$ and $k = 0.16$ as mentioned. We set $\rho = 0.1$, *i.e.* we accept extrema at least

¹<http://vision.deis.unibo.it/ssalti/Wave>

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

Algorithm 1 The WADE algorithm

Input: graylevel image I , threshold fraction $\rho \in [0, 1]$,
min and max features radius r_{min}, r_{max}

Output: A list \mathcal{K} of keypoints: (i_k, j_k, r_k)

```
1:  $N_0 \leftarrow \frac{r_{min}}{c\Delta t}, N \leftarrow \frac{r_{max}}{c\Delta t}, \mathcal{K} \leftarrow \emptyset$ 
2:
3:  $u^0 \leftarrow I$  ▷ initial conditions
4: for all inner pixels  $(i, j)$  do ▷ wave
5:    $u_{i,j}^{\frac{1}{2}} \leftarrow \text{eq. (5)}$ 
6: end for
7: for all boundary pixels  $(i, j)$  do
8:    $u_{i,j}^{\frac{1}{2}} \leftarrow \text{eq. (6)}$ 
9: end for
10: for all inner pixels  $(i, j)$  do ▷ heat
11:    $u_{i,j}^1 \leftarrow \text{eq. (8)}$ 
12: end for
13: for all boundary pixels  $(i, j)$  do
14:    $u_{i,j}^1 \leftarrow \text{eq. (6)}$ 
15: end for
16:
17: for  $n = 2 \rightarrow N - 1$  do
18:   for all inner pixels  $(i, j)$  do ▷ wave
19:      $u_{i,j}^{n+\frac{1}{2}} \leftarrow \text{eq. (4)}$ 
20:   end for
21:   for all boundary pixels  $(i, j)$  do
22:      $u_{i,j}^{n+\frac{1}{2}} \leftarrow \text{eq. (6)}$ 
23:   end for
24:   for all inner pixels  $(i, j)$  do ▷ heat
25:      $u_{i,j}^{n+1} \leftarrow \text{eq. (8)}$ 
26:   end for
27:   for all boundary pixels  $(i, j)$  do
28:      $u_{i,j}^{n+1} \leftarrow \text{eq. (6)}$ 
29:   end for
30: end for
31:
32: for  $n = N_0 \rightarrow N - 1$  do ▷ detection
33:    $r \leftarrow n * c * \Delta t, \Theta \leftarrow \rho * \Theta_{max}(r), t_1 \leftarrow t_1(r)$ 
34:   for all internal pixels  $(i, j)$  do
35:      $\bar{u}_{i,j}^n \leftarrow \text{eq (11)}$ 
36:     if  $u_{i,j}^n$  is max in  $3 \times 3 \times (1 + \frac{2}{c\Delta t})$  window then
37:       if  $|u_{i,j}^n - \bar{u}_{i,j}^n| > \Theta$  then
38:          $\mathcal{K} = \mathcal{K} \cup (i, j, r)$ 
39:       end if
40:     end if
41:   end for
42: end for
```

as sharp as 10% of the maximum sharpness θ_{max} , to allow low contrast symmetries to be detected in dark or overexposed images. We start to search for extrema after 12 iterations (*i.e.* the smallest feature we detect has radius $r_{min} = 6$

pixels) and we perform 200 iterations (*i.e.* the greatest feature has radius $r_{max} = 100$ pixels). In the legends, we refer to our proposal as WADE (Wave-based DETector).

5.1. Oxford Dataset

Fig. 7 reports results on the Oxford dataset. We present repeatability charts for each of the nuisances of the dataset. On the Boat set of images (Fig. 7a), which tests scale and rotation invariance, such theoretical properties of our algorithm are evidently confirmed. Moreover, the intuition that symmetries are a powerful and robust cue for finding repeatable keypoints is supported by the large margin in performance between our proposal and previous detectors based on other cues. Leuven images (Fig. 7b) allow for assessing robustness against lighting changes, and our method turns out on par with MSER, which is known to be the best detector for this nuisance [14]. On Graffiti images (Fig. 7c), which deal with robustness to affine deformations, our detector provides the same performance as 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 (Fig. 7d), the nuisance is JPEG compression: our detector turns out to be on par with the best previous detectors, Hessian-Laplace and Harris-Laplace. Finally, Bikes images (Fig. 7e) show how symmetries are definitely the best cue to overcome blur: this makes sense as, intuitively, symmetries are largely unaffected by isotropic smoothing.

As for the number of correspondences, we cannot report the charts here due to space limits. However, our detector finds a number of keypoints similar to DoG, which is usually slightly larger than the other methods: for example, the scene with more keypoints is Boat where both methods find about 7000 keypoints on the reference image, whereas the scene with less features is Bikes, where our method finds about 650 keypoints and DoG about 750 keypoints.

5.2. Robot Dataset

Fig. 7f shows the results for the Robot Dataset. Here we focus on the linear path, which tests scale invariance, with stable, diffuse illumination (the experiment depicted by Fig. 9.d in [1]). 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 to speed up the tests. Hence results for our method are available only at the minimum, middle and maximum scale variation, whereas results for all other detectors were kindly provided by the authors of [1]. The proposed algorithm shows the best recall rate at the middle 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 dataset and reinforces them by demonstrating that they hold for a variety of scenes.

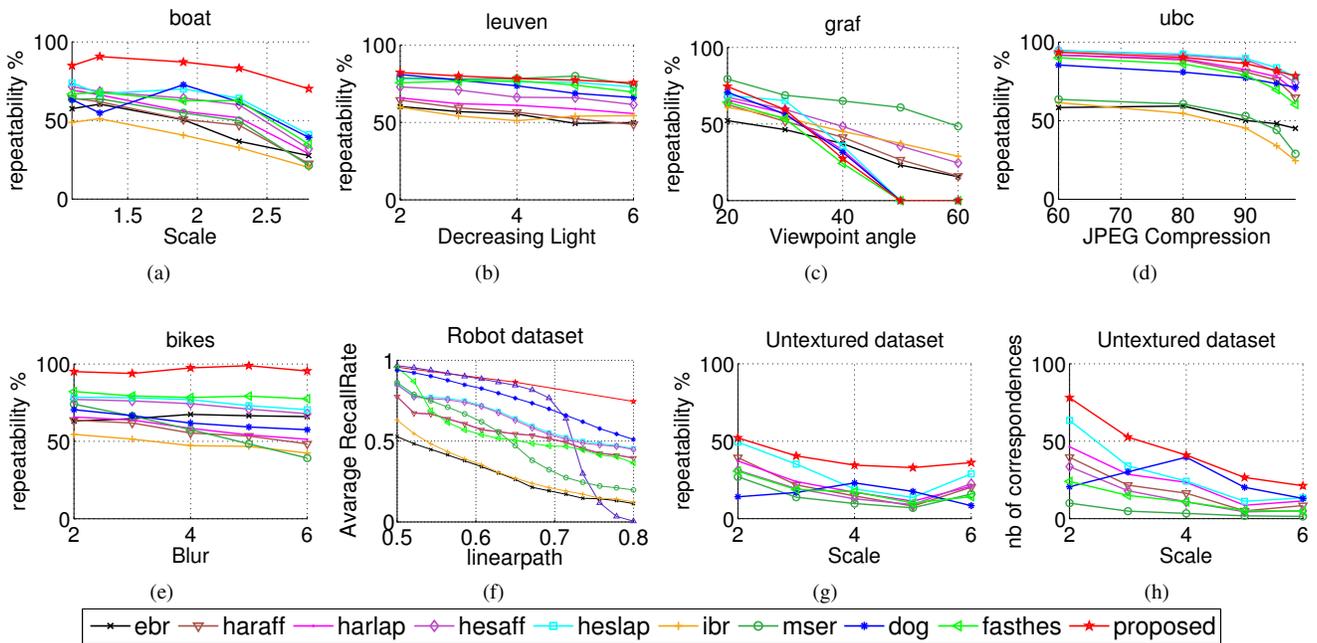


Figure 7: a)-e) Repeatability on 5 sets of images of the Oxford dataset. f) Mean recall rates on all the 60 sets of images of the Robot dataset along the linear path. g-h) Mean repeatability and number of correspondences on the 5 sets of images of the Untextured dataset. The x axis represents the level of difficulty of the considered nuisance. In (f) it reports the distance from the scene in metric units.

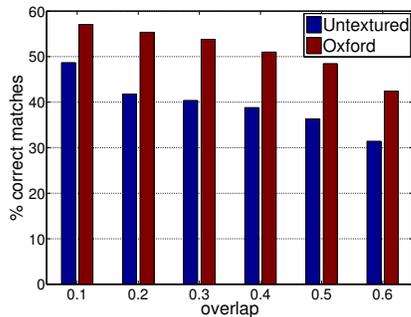


Figure 8: Correct matches of SIFT descriptors versus overlap error of keypoints on Oxford and untextured datasets.

5.3. Untextured Dataset

Fig. 7g-7h report the results on our novel, untextured objects dataset. We provide the mean repeatability and mean number of correspondences for each method. 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 (Fig. 7h) by two orders of magnitude with respect to the other

datasets, and repeatability falls below 50%. Please note that on this dataset we lower from 40% to 20% the overlap error to consider two keypoints as repeatable: as shown by Fig. 8, we found that on the Oxford dataset, a robust descriptor (*i.e.* SIFT) can correctly match about 51% of the keypoints whose overlap error is smaller than 40% but on our tougher dataset only 38% of such keypoints can be matched, and 48% can be achieved only with overlap errors smaller than 10%. Therefore, we found that with untextured objects keypoints should be localized more accurately, when used with existing descriptors, to reach performance comparable to those attained on planar, highly textured objects. Hence, we selected 20% as a guideline. 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 presence of occlusions.

6. Concluding remarks

We have shown that the traditional linear scale-space theory may not be the unique way to address scale-invariant feature detection within an elegant conceptual framework. Instead of the heat equation, hyperbolic PDEs, and in par-

ticular 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 these to provide robust and discriminative cues for detection of repeatable keypoints on a broader set of image structures than those found on planar highly textured surfaces.

As shown by prominent evaluations [1, 14], salient regions are usually complementary, so that a bunch of diverse detectors might be deployed whenever one cannot predict in advance the likely scene structures. The proposed algorithm, based on an underexploited cue as symmetry, might turn out highly complementary with respect to current state-of-the-art detectors. As vouched by earlier work [4], our approach is nicely implementable on fine grained data-parallel architectures, and as such holds the potential for effective acceleration via GPU processing.

We hope that the encouraging results shown in this paper will generate interest toward our intuition, as many questions remain unanswered. First of all, a theoretical investigation should be carried out to understand the relationship between the traditional Gaussian scale-space and our wave-based family of signals. Next, more sophisticated numerical schemes for discretizing the wave equation for feature detection might be helpful: in particular, to reduce numerical dispersion and diffusion as well as to achieve higher rotational invariance. Finally, several interesting extensions are possible, such as defining a symmetry descriptor from $u(x, y; t)$ and extending feature detection to color images via the model of images as manifold embedded in a combined spatial-color space [20].

References

- [1] H. Aanæs, A. L. Dahl, and K. Steenstrup Pedersen. Interesting interest points. *Int. J. Comput. Vision*, 97(1):18–35, mar 2012. 5, 6, 8
- [2] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (SURF). *Comput. Vis. Image Underst.*, 110(3):346–359, jun 2008. 5
- [3] R. H. Chan, A. Lanza, S. Morigi, and F. Sgallari. An adaptive strategy for the restoration of textured images using fractional order regularization. *Numerical Mathematics: Theory, Methods and Applications*, 6(1):276–296, jan 2013. 1
- [4] 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. 1, 8
- [5] B. Engquist and A. Majda. Absorbing boundary conditions for the numerical simulation of waves. *Mathematics of Computation*, 31(139):629–651, jul 1977. 2
- [6] K. Hanahara and M. Hiyane. A circle-detection algorithm simulating wave propagation. *Mach. Vision Appl.*, 4(2):97–111, mar 1991. 1, 3
- [7] D. Hauage and N. Snavely. Image matching using local symmetry features. In *Proc. IEEE Conf. Computer Vision Pattern Recognition*, pages 206–213, 2012. 2
- [8] J. Koenderink. The structure of images. *Biological Cybernetics*, 50(5):363–370, 1984. 1
- [9] T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, Norwell, MA, USA, 1994. 1
- [10] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, nov 2004. 1, 5
- [11] G. Loy and A. Zelinsky. Fast radial symmetry for detecting points of interest. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(8):959–973, aug 2003. 2
- [12] 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(12):1579–1590, dec 2003. 1
- [13] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *Proc. British Machine Vision Conference*, volume 1 of *BMVC’02*, pages 384–393, 2002. 5
- [14] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool. A comparison of affine region detectors. *Int. J. Comput. Vision*, 65(1-2):43–72, nov 2005. 5, 6, 8
- [15] S. Osher and L. I. Rudin. Feature-oriented image enhancement using shock filters. *SIAM J. Numer. Anal.*, 27(4):919–940, aug 1990. 1
- [16] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(7):629–639, jul 1990. 1
- [17] V. Ratner and Y. Y. Zeevi. Denoising-enhancing images on elastic manifolds. *Trans. Img. Proc.*, 20(8):2099–2109, aug 2011. 1
- [18] D. Reissfeld, H. Wolfson, and Y. Yeshurun. Context-free attentional operators: the generalized symmetry transform. *Int. J. Comput. Vision*, 14(2):119–130, mar 1995. 2
- [19] G. L. Scott, S. C. Turner, and A. Zisserman. Using a mixed wave/diffusion process to elicit the symmetry set. *Image Vision Comput.*, 7(1):63–70, feb 1989. 1, 3
- [20] N. Sochen and Y. Y. Zeevi. Images as manifolds embedded in spatial-feature non-euclidean space. Technical Report CCIT260, Israel Inst. Technol, Haifa, Haifa, Israel, 1998. 8
- [21] H. Tek and B. B. Kimia. Symmetry maps of free-form curve segments via wave propagation. *Int. J. Comput. Vision*, 54(1-3):35–81, aug 2003. 1, 3
- [22] T. Tuytelaars and L. Van Gool. Matching widely separated views based on affine invariant regions. *Int. J. Comput. Vision*, 59(1):61–85, aug 2004. 5
- [23] 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)*, pages 205–212, 2008. 3
- [24] A. P. Witkin. Scale-space filtering. In *Proc. Eighth Int. Joint Conf. on Artificial Intelligence*, volume 2 of *IJCAI’83*, pages 1019–1022, 1983. 1