

The Improbability of Harris Interest Points

Marco Loog and François Lauze

Abstract—An elementary characterization of the map underlying Harris corners, also known as Harris interest points or key points, is provided. Two principal and basic assumptions made are: 1) Local image structure is captured in an uncommitted way, simply using weighted raw image values around every image location to describe the local image information, and 2) the lower the probability of observing the image structure present in a particular point, the more salient, or interesting, this position is, i.e., saliency is related to how uncommon it is to see a certain image structure, how surprising it is. Through the latter assumption, the axiomatization proposed makes a sound link between image saliency in computer vision on the one hand and, on the other, computational models of preattentive human visual perception, where exactly the same definition of saliency has been proposed. Because of this link, the characterization provides a compelling case in favor of Harris interest points over other approaches.

Index Terms—Interest points, saliency, Harris corners, visual attention, low probability, elementary characterization.

1 INTRODUCTION

IMAGE matching, image classification, geometric hashing, image registration, and many other computer vision and image analysis techniques rely on and benefit from the possibility to find salient or interest points in a scene [1], [2], [3]. Such points are assumed to capture conspicuous structures in images that are descriptive enough to solve the task at hand, thus avoiding the processing of the entire image and allowing for rapid, potentially real-time solutions to various computer vision and image analysis tasks.

Many different approaches to determining such interest points have been proposed for which a couple of different underlying principles have been suggested that, supposedly, drive the interestingness of image locations. Starting from the work by Förstner and Gülch [4] and Harris and Stephens [5], Triggs [6] casts key-point detection in a rather nice and general framework that relies on stability as the defining concept. At the basis of some earlier approaches, there have been energy and information theoretic considerations [3], [7], but also novelty and outlier detection techniques have been studied as a way to formalize point saliency [8], [9]. Most of the considerations are of a (differential) geometric nature [1], [10], [11] and often focus on a description of actual corners or vertices in the image [12], [13]. One should realize, however, that these latter operators are not very specific and respond to more general image structures.

The same goes for Harris interest points [2], [5], [6]. Originally, the underlying map was considered as a measure of cornerness and the points detected based on it have been referred to as Harris, or Plessey, corners, but they are generally considered to indicate how

- M. Loog is with the Pattern Recognition Laboratory, Delft University of Technology, HB 11.070, Mekelweg 4, 2628 CD Delft, The Netherlands, and The Image Group, Department of Computer Science, University of Copenhagen, Universitetsparken 5, 2100 Kbh Ø, Copenhagen, Denmark. E-mail: m.loog@tudelft.nl.
- F.B. Lauze is with The Image Group, Department of Computer Science, University of Copenhagen, Universitetsparken 5, 2100 Kbh Ø, Copenhagen, Denmark. E-mail: francois@diku.dk.

Manuscript received 29 Sept. 2009; revised 24 Dec. 2009; accepted 12 Jan. 2010; published online 25 Feb. 2010.

Recommended for acceptance by S. Belongie.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2009-09-0650.

Digital Object Identifier no. 10.1109/TPAMI.2010.53.

much a location stands out in an image and they do not necessarily refer to features readily identifiable as corners or blobs [6] (see also [14]). Further on in this paper, an illustration using the image in Fig. 1, showing one of the well-known pop-out examples, provides an, initially possibly surprising, additional example of the fact that Harris is not only sensitive to corners or corner-like structures.

Harris interest point detection is a very popular scheme to find salient points and many methods rely on the original formalism or at least take the approach as a basis for further developments [14], [15], [16], [17]. Overall, Harris interest points have been shown to work relatively well in comparative studies involving several other schemes.

1.1 Contribution and Suppositions

This paper shows that in an unsupervised, task-independent, and uncommitted setting, plausible requirements can be imposed on the interest point detection task that uniquely set apart the Harris interest map from all other possible choices. It provides an elementary characterization of the underlying map and allows for further generalizations of the Harris interest point detector in a principled way. The main assumptions made are, first, the lower the probability of occurrence of image structure associated to a certain image location is, the more interesting or salient this location is (cf. [8], [9]) and, second, local image structure is represented in an uncommitted way, merely using weighted raw image gray values.

The first assumption in our characterization directly relates our approach to particular task-independent, low-level theories of computational, preattentive visual attention, which in various settings have been demonstrated to correlate well to specific eye tracking data (see [18], [19], [20], [21], [22]). The precise computational definition of saliency we rely on has been studied before in the visual perception literature, in [18], [22], for instance. We see this link with biological vision as part of a tradition in certain areas of low-level computer vision [1], [11], [23], [24] and a point of interest in its own respect.

The second assumption allows us to find a closed-form solution for the density function over the distribution of all patches from a single image. This finding, in combination with the first supposition, then permits us to derive a closed-form expression of saliency as well. Finally, the latter expression turns out to be equivalent to the original Harris interest map.

1.2 More on Related Work

As pointed out in the previous section, within the field of computational visual attention, various attempts have been made to provide a basic characterization of image saliency. These characterizations are typically operationalizations of low-level and preattentive saliency as more qualitatively described in, for instance, [21] and [19]. The earlier mentioned [18] provides a good example of this approach. It defines image locations to be salient if the information content of a single observation is high, i.e., if minus the logarithm of the probability is high. This, indeed, is equivalent to some early approaches from the computer vision literature [8], [9] which aim to find and exploit points with low probability.

An interest point detector that also relies on an information theoretic formulation is presented in [25]. In this case, however, the local entropy of the gray values is considered as a measure of saliency and not, like [18] and related approaches, the entropy of a certain location in the context of the other locations.

Further works from the area of visual perception rely on a similar, if not exactly the same, operationalization as the one we exploit [20], [22], cf. [26]. A slightly more elaborate and involved scheme has been presented recently in [27] and an early reference from this community is [28]. The latter provides a discriminative instead of a more information theoretic approach but is still rather

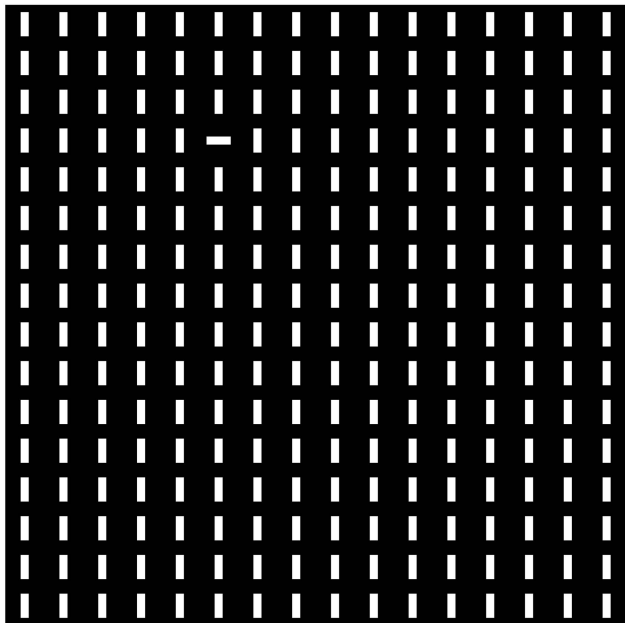


Fig. 1. Gray-value image on which some points in our approach are exemplified.

directly related to previous works mentioned. A further reference is [29], which also takes a low probability point of view to saliency, defining it to be proportional to one over the probability of observing certain local features in a context, which therefore relates directly to our approach as well. The definition of saliency we rely on has been validated using actual human eye tracking data. We should mention that various other proposals to quantifying low-level image saliency exist that may not be directly related to the one of interest in this work (see, for instance, [30], [31], [32]).

1.3 Outline

Section 3 formulates the aforementioned requirements in a more precise way and formulates the main result in Section 3.3. To start with, the same section introduces some necessary notation. Prior to this, Section 2 briefly recapitulates Harris interest point detection with a focus on the underlying map. Section 4 concludes the paper and offers a discussion. For the sake of rigor, the proof of the principal result from Section 3 is provided in the Appendix.

2 HARRIS INTEREST POINTS

Harris interest point detectors [2], [5] and its variations [4], [6], [14], [16] are typical examples of the interest point detection techniques currently employed in many computer vision methods and applications, reaching even the surface of Mars [33].

For a two-dimensional image $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, the saliency map underlying Harris interest points, denoted H , measures the local two-dimensional gray-value variability in every location x . This is quantified by means of the structure tensor $T(x) = (k * (\nabla f \nabla f^t))(x)$ and is taken to equal $H(x) = \det T(x) - \kappa \text{trace}^2 T(x)$, in which κ is a small nonnegative constant which artificially suppresses unwanted response to salient edge locations. In this paper, we merely deal with the case for which κ equals zero.

The convolution kernel k is applied entry-wise to the outer product of the gradient. It is typically taken to be Gaussian. Given the mapping, the interest points are normally taken from its local maxima only, which models an inhibition of return mechanism (cf. [21]) and prevents locations in the direct vicinity of the maxima from being chosen as well. Our initial interest goes out to the map

H for general single-valued images from \mathbb{R}^n to \mathbb{R} , though Section 4 discusses some further generalizations.

3 HARRIS AS VISUAL SALIENCY

To start with, we need to impose mild conditions on the images f we consider. In fact, we will need a weak form of differentiability and, therefore, assume $f \in W^{1,2}(\mathbb{R}^n)$, the Sobolev space of square-integrable functions with square-integrable partial (weak) derivatives [34]. This choice of function space is not a severe restriction as, generally, operations cannot be applied directly to a raw image, but typically are done on the image observed through a smooth enough aperture, which results in a smooth enough image [35].

3.1 Image Structure

Second, we need to define how image structure is captured and associated to every location in an image f . (Fig. 1 displays the example image used throughout the remainder.) Generally, the local image content in the image f is described by means of a collection of features, e.g., outputs of linear filters or more complex functions of the image f . Here, we aim at an uncommitted description of image structure and avoid the use of specialized and dedicated filter outputs to describe local image information. This leads us to simply represent every image location x by the full raw image modulated by a localized weighting function ω (like, for example, in Fig. 2a), i.e., every location is represented by an image patch as illustrated in Figs. 2b, 2c, and 2d.

To formalize this, we follow the notations from [36] and define the translation operator τ_x by $(\tau_x f)(y) = f(y - x)$ and mirroring operator \check{f} by $\check{f}(x) = f(-x)$. We can now express the map $\Phi_f: \mathbb{R}^n \rightarrow L^2(\mathbb{R}^n)$, where $L^2(\mathbb{R}^n)$ is the space of square-integrable functions on \mathbb{R}^n , which relates every image position in \mathbb{R}^n to a weighted image patch, as

$$\begin{aligned} \Phi_f: x &\mapsto \omega \cdot \tau_x \check{f} \\ &= \{a \mapsto \omega(a) \tau_x \check{f}(a) = \omega(a) f(x - a)\}, \end{aligned} \quad (1)$$

where ω is bounded, i.e., ω is in $L^\infty(\mathbb{R}^n)$.

Typically, ω is chosen rotationally invariant because of the a priori absence of any directional preference and the value $\omega(x)$ would decrease with the distance of x from the origin. The latter reflects that image intensities in the vicinity of the image location under scrutiny are more important than more distant intensities. It may be considered a rough way of dealing with loss of acuity with increase of eccentricity.

3.2 Quantification of Saliency

The next step is to make explicit the definition of low-level saliency employed by us and previously used in [8], [9], [18], [22], [29]. For this, define $\mathcal{S}_f := \Phi_f(\mathbb{R}^n)$ to be the set of all raw image patches and denote a possible probability density on this domain by π . A location x in an image is more salient than another location y if and only if the probability of the image patch $\Phi_f(x)$ associated to x is less probable than patch $\Phi_f(y)$ which is associated to location y , i.e., if and only if $\pi(\Phi_f(x)) < \pi(\Phi_f(y))$. Now the next section determines the particular form the density π takes under the mapping Φ_f , leading to the principal result of the paper.

Before coming to this next section, we point out that our derivation deals in principle with a single image. The same relation between Harris and saliency, however, holds in case we consider a finite set of generic images, $\{f_1, f_2, \dots\}$. For such a generic set, it holds that all sets of raw image patches \mathcal{S}_{f_i} are disjoint in patch space and so all images can be treated separately, irrespective of the other images in the set.

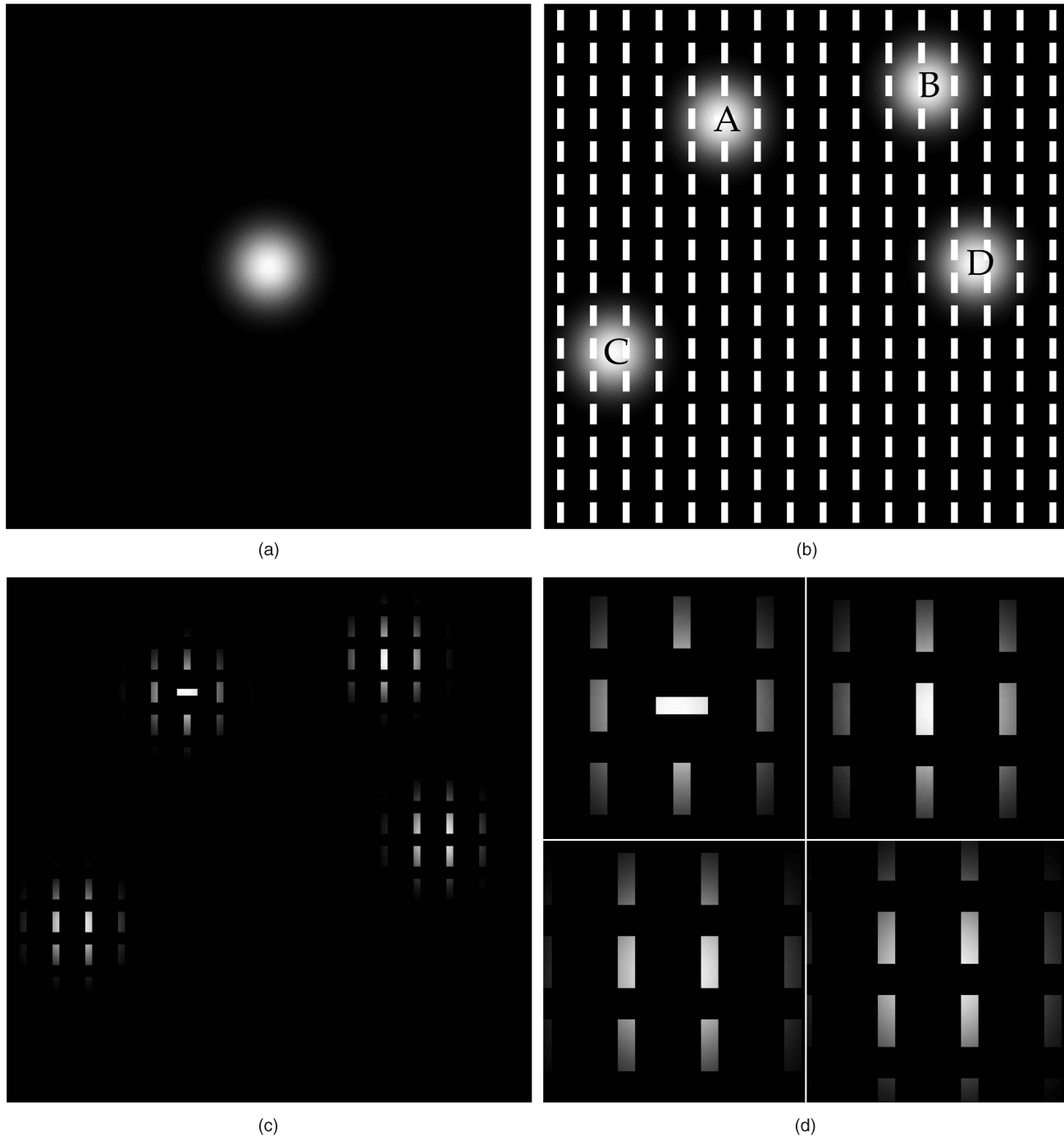


Fig. 2. Representing local image structure by patches: (a) an example of a localized weighting function ω ; (b) example images locations A, B, C, and D from which patches are extracted; (c) modulations of original image, $\omega \cdot \tau_x f$, for the four locations (displayed in a single image); (d) enlarged versions of the four final example patches.

3.3 Main Results

The collection of patches gives rise to a density on $S_f \subset L^2(\mathbb{R}^n)$ just as much as feature vectors in a feature space induce a density according to their distribution in that space. Now, the following theorem makes explicit how the raw image patches are distributed when we know the probability density function p on the image domain \mathbb{R}^n . It is noted that, in contrast to previous work, the density is not estimated but can actually be calculated in closed form in this case.

Theorem 1. Assume $\omega > 0$. Then, a density p on the image domain \mathbb{R}^n , with respect to the Lebesgue measure of \mathbb{R}^n , induces, under the

mapping Φ_f , a density π on S_f , with respect to the measure induced by the L^2 -metric, given by

$$\pi = \frac{p}{\sqrt{H}} \circ \Phi_f^{-1}, \tag{2}$$

where the Harris map H is the determinant of the structure tensor T obtained by convolving $\nabla f \nabla f^t$ with the kernel ω^2 .

Now, what we are chiefly interested in is which patch probability is associated to every location, as this is, by definition, inversely relates to its saliency. The probability we are looking for

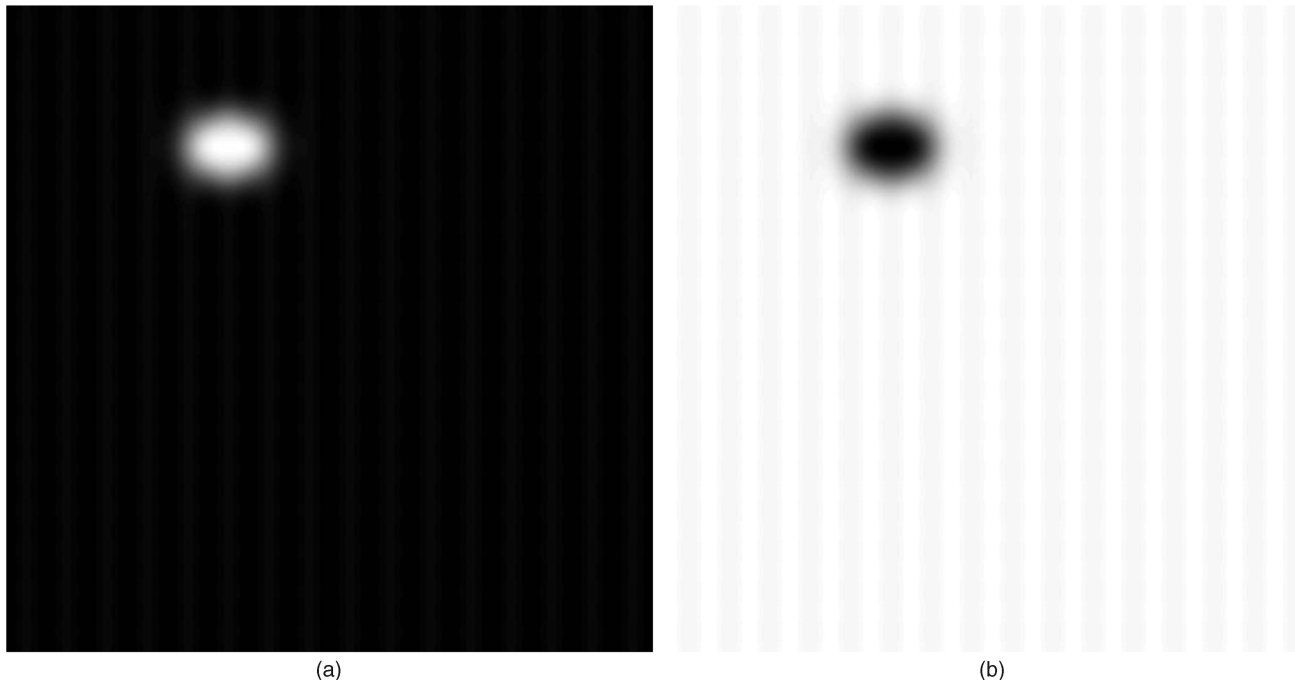


Fig. 3. Quantities related to saliency for Fig. 1: (a) displays the Harris map H and (b) the logarithm of the probabilities of every patch associated to every image position. As shown (see main text), (a) and (b) are related based on $1/\sqrt{H} \propto \pi \circ \Phi_f$. In (b) one is able to see not only the location of the pop-out, but also the saliency response caused by the dashed vertical lines against the black background.

equals $\pi(\Phi_f(x))$ for every image location x , i.e., $\pi \circ \Phi_f$ is the probability density of interest. Now, mainly rewriting the result from Theorem 1 leads in fact to the principal result of our paper.

Corollary 1. *Provided no image location is a priori preferred over any other, meaning p is chosen to be uniform, the Harris map H is inversely proportional to the square of $\pi \circ \Phi_f$.*

Or, in other words, the Harris map provides a measure for the saliency of particular image locations, as the latter is inversely proportional to the probability of observing a specific image structure. As an illustration, Fig. 3a shows the Harris map H for the image in Fig. 1, while Fig. 3b displays the log probability ($\log \circ \pi \circ \Phi_f$) for every image location of the same image. The location of the pop-out in Fig. 1 is clearly visible in both Figs. 3a and 3b. The latter also shows, though rather vaguely, an increase in saliency of the vertical and dashed white lines against the black background. Their saliency, however, does not compare to the single white horizontal stroke. In addition, Figs. 1, 2, and 3 nicely illustrate that Harris maps indeed respond to certain structures that are not corner-like.

4 DISCUSSION AND CONCLUSION

An elementary, probabilistic characterization of the map underlying Harris interest points has been given. One appealing aspect of our characterization is that it connects this well-known operator from computer vision to some of the probabilistic approaches to computational visual attention as currently emerging in the literature on preattentive visual perception. Another attractive feature is that the theory, while making minimal assumptions, provides one of the few generic, basic, and theoretical considerations that may lead one to favor one interest operator over the other, making a compelling case in favor of Harris key points.

A core assumption in our characterization is, first of all, that image structure is represented in an uncommitted way and can simply be taken to be a weighted image patch $\omega \cdot \tau_x \cdot f$ for every image location x . For this to hold, the convolution kernel k

employed in Harris interest point detection, should be chosen equal to the square of this nonnegative weighting function ω . In a way, this choice of straightforward features avoids a bias to measuring particular, dedicated, and involved image structures. Choosing more complicated, nonlinear, local image descriptors may certainly be desirable in particular situations (see below), but in these cases we cannot link saliency directly to Harris anymore (not to mention that our current derivation possibly cannot be used directly anyway). Thus, our choice of image descriptors is essential and it is precisely the uncommitted choice employed in this work that leads to the Harris interest map.

Certain types of weighting schemes can be justified to some extent. For instance, an isotropic kernel ω , whose value drops off with the distance to its origin reflects the fact that close by values in a scene are likely to carry more information about the central location than remote positions. Arguments that would justify one and only one specific shape are, however, lacking thus far. Nonetheless, one may suspect that a Gaussian weighting is the favorite candidate (in this case, the kernel $k = \omega^2$ in the Harris map would also be Gaussian).

The other principal postulate is that the low-level, task-independent saliency of an image location is inversely related to the probability associated to the image structure present. In a way, this makes explicit that the interestingness of an observation is related to how often such a configuration is observed, how uncommon it is, how surprising or unlikely. It is also a principled measure used to decide, for instance, on the novelty or outlieriness of an observation [37]. This requirement, we therefore believe, is a natural one, and in the design of different saliency maps this may be the principle that remains unaltered. In that case, the idea would be that with additional knowledge about the key-point detection task, it should be the image descriptor that is changed.

For example, knowing that changes in absolute luminance should not alter the saliency measure, such additional knowledge could be integrated by making the image patch representation more committed and turning $\omega \cdot \tau_x \cdot f$, from (1), into a luminance-invariant version in some manner. Obviously, various ways to

enforce luminance-invariance can be thought of and additional arguments might be needed to prefer one approach over the other. Another relevant, yet possibly nontrivial, extension is to use our scheme for key-point detection not only over spatial location but also over the so-called local scale and integration scale of the Harris operator, a topic that has been considered previously in, for instance, [14] and [38].

There are two examples from the literature already, in which our approach would provide a similar elementary characterization for the saliency map suggested. With minor alterations to the setting described in this paper (considering a spatiotemporal weighting function ω and multichannel images, respectively), space-time interest operators [16] and Harris color operators [17] can be derived, i.e., two instances that indeed illustrate the broader applicability of the approach presented. Another one is defining a principled three-dimensional interest operator, which, not surprisingly, would simply be the determinant of the three-dimensional structure tensor, reminiscent of what can already be found in the literature [39].

All in all, this work rigorously demonstrates the elementary fact that a perception-derived, low-level, probabilistic criterion together with an uncommitted approach to describing local scene or image content characterizes the saliency map underlying one of the most used interest point detectors within computer vision, i.e., Harris interest point detector.

Let us finally note that the definition of visual saliency employed in this work is indeed task independent and low level and for particular computer vision tasks, potentially more powerful interest point detectors might be constructed on the basis of more high-level computational theories of visual search and attention. These theories extend the approaches mentioned in Section 1.2, combine top-down, task-dependent, and bottom-up mechanisms, and go beyond preattentive vision (see, for instance, [21], [29], or [40], [41], and possibly [42] for a more general exposition). These ideas, however, have not yet crystallized as much as is the case for preattentive vision. Nonetheless, an attentive, top-down, and high-level Harris map may be imminent.

APPENDIX A

MAIN PROOF

We here provide a rigorous proof of Theorem 1. The proof relies on elementary functional analysis, but may not be of interest in every single detail to all readers. Therefore, we first give the following gist of the proof.

Assume that the map $\Phi_f : \mathbb{R}^n \rightarrow L^2(\mathbb{R}^n)$ is a C^1 -embedding and let $\mathcal{S}_f = \Phi_f(\mathbb{R}^n)$ be the set of raw image patches. In this case, this set \mathcal{S}_f carries the structure of a submanifold of $L^2(\mathbb{R}^n)$ and Φ_f becomes a global parametrization of \mathcal{S}_f , i.e., a diffeomorphism between \mathbb{R}^n and \mathcal{S}_f . The restriction of the standard scalar product of $L^2(\mathbb{R}^n)$ given by $\langle f, g \rangle_2 = \int_{\mathbb{R}^n} fg \, dx$, endows \mathcal{S}_f with a structure of Riemannian manifold and the expression g^f of the Riemannian metric in the global coordinate system \mathbb{R}^n provided by Φ_f is the structure tensor T and H is precisely the Gram determinant $\det(g_{ij}^f)$ of g^f . From this, Theorem 1 and its related corollary follow easily, using the change of variables theorem for manifold integral calculus [43].

In fact, this is a situation similar to the one encountered in standard probability theory: Φ_f would map one vector of random variables to another and, under certain conditions, the probability density of the one can be expressed in terms of the other using the change of variable theorem. The condition we need to be fulfilled in our setting is that Φ_f is a differentiable embedding. The main point is to prove that this is indeed the case.

The actual proof of Theorem 1 is thus split up in two sections. First, we establish in Proposition 1 the link with the structure tensor and the measure theoretic results from which Theorem 1

follows. The second section then considers the necessary requirement for Φ_f to be an embedding.

A.1 From Embedding to Structure Tensor

Proposition 1. Assume that Φ_f is an embedding.

1. The metric g^f on \mathbb{R}^n induced by the standard Hilbertian L^2 -metric on $\mathcal{S}_f := \Phi_f(\mathbb{R}^n)$ is the ω^2 -structure tensor of f :

$$g_{ij}^f(x) = \omega^2 * (f_{x_i} f_{x_j})(x),$$

that is,

$$g^f(x) = \omega^2 * (\nabla f \nabla f^t)(x).$$

2. The canonical measure associated to the Riemannian structure of \mathcal{S}_f is, when expressed in the global coordinate system Φ_f , given by

$$\mu_f = \sqrt{\det(g_{ij}^f)} \lambda_n,$$

where λ_n is the Lebesgue measure on \mathbb{R}^n and $\det(g_{ij}^f)$ is the Harris Map H .

For $x \in \mathbb{R}^n$, $y = \Phi_f(x)$, we denote by $T_x \mathbb{R}^n$ the tangent space of \mathbb{R}^n at x (one has of course $T_x \mathbb{R}^n \equiv \mathbb{R}^n$), $T_y \mathcal{S}_f$ the tangent space of \mathcal{S}_f at y and $d_x \Phi_f$ the differential map $T_x \mathbb{R}^n \rightarrow T_y \mathcal{S}_f$ of Φ_f at x . In the standard coordinate system of \mathbb{R}^n , this is the matrix of partial derivatives of Φ_f and a direct calculation shows that $\frac{\partial \Phi_j}{\partial x_i}(x) \stackrel{\text{def}}{=} d_x \Phi_j e_i = \Phi_{f_{x_i}}(x) = \omega \tau_x f_{x_i}$ where f_{x_i} is the partial derivative of f with respect to x_i , and by definition of g_{ij}^f ,

$$\begin{aligned} g_{ij}^f(x) &= \langle \omega \tau_x f_{x_i}, \omega \tau_x f_{x_j} \rangle_2 \\ &= \int_{\mathbb{R}^n} \omega^2(a) f_{x_i}(x-a) f_{x_j}(x-a) \, da \\ &= \omega^2 * (f_{x_i} f_{x_j})(x), \end{aligned}$$

which establishes the result in point 1. The second point is a classical and elementary result from Riemannian geometry, see [44], for instance.

Theorem 1 then follows immediately from point 2 of the above proposition. In order to finish the proof, we thus need to prove that if ω remains strictly positive, then Φ_f is indeed a differential embedding as soon as f carries some information, i.e., $f \neq 0 \in W^{1,2}(\mathbb{R}^n)$, and this is established in the following theorem:

Theorem 2. Assume $\omega \in L^\infty(\mathbb{R}^n)$ and is strictly positive and $f \in W^{1,2}(\mathbb{R}^n)$ is nonzero, then Φ_f is a C^1 -embedding of \mathbb{R}^n into $L^2(\mathbb{R}^n)$.

A.2 Proof of Theorem 2

We start by describing the different, unsurprising steps:

1. Φ_f is continuously differentiable,
2. Φ_f is an immersion,
3. Φ_f is into, and
4. Φ_f is an homeomorphism onto its image.

Before proving them, we first notice that the mirroring operation $f \mapsto \tilde{f}$ is linear, continuous, and in fact a self-inverse isometry of $L^2(\mathbb{R}^n)$. The translation τ_x is also an isometry, with inverse τ_{-x} , and we have the following useful result:

Lemma 1. Let $f \in L^2(\mathbb{R}^n)$ and $x \in \mathbb{R}^n$.

1. If $f = \tau_x f$, then either $x = 0$ or $f = 0$ a.e.,
2. If $f = -\tau_x f$, then $f = 0$ a.e.

For point 1: Assume $x \neq 0$. Then, for all $k \in \mathbb{Z}$, $f(a) = f(a + kx)$ a.e., but $f(a + kx) \rightarrow 0$ a.e. when $|k| \rightarrow \infty$ since $\|a + kx\| \rightarrow \infty$ and

$f \in L^2(\mathbb{R}^n)$. Therefore, $f = 0$ a.e. The proof of point 2 is similar, apart from the case $x = 0$, which trivially implies that $f = 0$ a.e.

The continuous differentiability of Φ_f will follow from the more general lemma stated below.

Lemma 2. *Let $g \in L^2(\mathbb{R}^n)$. Then, Φ_g is continuous. Let $x, y \in \mathbb{R}^n$. Then,*

$$\|\Phi_g(x) - \Phi_g(y)\|_2^2 = \int_{\mathbb{R}^n} \omega^2 (\tau_x \check{g} - \tau_y \check{g})^2 dx \quad (3)$$

$$\leq \|\omega^2\|_\infty \|\tau_x \check{g} - \tau_y \check{g}\|_2^2$$

$$\text{(by Hölder inequality)} \quad (4)$$

$$= \|\omega^2\|_\infty \|\check{g} - \tau_z \check{g}\|_2^2$$

$$\text{(with } z = y - x) \quad (5)$$

and, since mirroring is an isometry, it is enough to show that $z \mapsto \tau_z g$ is continuous; a classical result shows that it is in fact uniformly continuous for $g \in L^p(\mathbb{R}^n)$, $p \geq 1$ (see, for instance, [45]).

We mentioned already that $\partial \Phi_f / \partial x_i = \Phi_{f_{x_i}}$, where f_{x_i} is the partial derivative of f with respect to x_i . So, in order to show that Φ_f is \mathcal{C}^1 , we apply the previous lemma to the f_{x_i} s, they are in $L^2(\mathbb{R}^n)$ since $f \in W^{1,2}(\mathbb{R}^n)$. Its partial derivatives being continuous, Φ_f is \mathcal{C}^1 .

Next, we come to show that Φ_f is an immersion, i.e., that for each $x \in \mathbb{R}^n$, the differential $d_x \Phi_f$ of Φ_f at x is into. Pick an $x \in \mathbb{R}^n$ and $v = (v_1, \dots, v_n) \in \mathbb{R}^n$ such that $d_x \Phi_f v = 0$. $d_x \Phi_{f_{x_i}} v$ is the function $\sum_{i=1}^n v_i \omega \tau_x \check{f}_{x_i}$. Assuming that this function is $0 \in L^2(\mathbb{R}^n)$ and since $\omega > 0$, we obtain

$$0 = \sum_{i=1}^n v_i \tau_x \check{f}_{x_i} = \tau_x \left(\sum_{i=1}^n v_i \check{f}_{x_i} \right) = -\tau_x \left(\sum_{i=1}^n v_i f_{x_i} \right), \quad (6)$$

and thus $\sum_{i=1}^n v_i f_{x_i} = \nabla f \cdot v = 0 \in L^2(\mathbb{R}^n)$ since τ_x is an isometry. If $v \neq 0$, then one can assume without loss of generality that $v = (1, 0, \dots, 0)$ and $f_{x_1} = 0 \in L^2(\mathbb{R}^n)$ and thus f is constant along lines parallel to the x_1 axis. It follows that $f = 0$ a.e. in $L^2(\mathbb{R}^n)$.

In the third step, we prove that Φ_f is into. First, given $s \in \mathcal{S}_f$, we compute, for $x \in \mathbb{R}^n$, the "cosine" map

$$c_s(x) = \frac{\langle s, \Phi_f(x) \rangle_2}{\|s\|_2 \|\Phi_f(x)\|_2}.$$

Then, we claim that, for all $s \in \mathcal{S}_f$, c_s has a unique maximizer $x(s)$ that satisfies $\Phi_f(x(s)) = s$, $x(\Phi_f(x)) = x$, from which it follows that Φ_f is into.

From the Cauchy-Schwarz inequality, $c_s(x) \leq 1$ and if $s = \Phi_f(y)$, this maximum is reached at $x = y$. Let $\bar{x} \in \mathbb{R}^n$ maximizing c_s , i.e., $c_s(\bar{x}) = 1$. Then, once again from the Cauchy-Schwarz inequality, $\Phi_f(\bar{x})$ and $\Phi_f(y) = s$ are linearly dependent: $\exists \lambda \neq 0$, $\lambda \omega \tau_{\bar{x}} \check{f} = \omega \tau_y \check{f}$, and since $\omega > 0$ and mirroring is an isometry, we have $\lambda \tau_{\bar{x}} f = \tau_y f$. Since $\tau_{\bar{x}}$ and τ_y are isometries of $L^2(\mathbb{R}^n)$, taking the L^2 -norm on both sides implies $|\lambda| = 1$, i.e., one has $\tau_{y-\bar{x}} f = \pm f$. Then, $\lambda = 1$ and $\bar{x} = y$ by Lemma 1.

Because \mathbb{R}^n is not compact, it is not enough to check that Φ_f is into in order to end the proof of the theorem. We need to check that Φ_f is an homeomorphism onto its image, and we do that by directly showing that the inverse map $\phi_f : \mathcal{S}_f \rightarrow \mathbb{R}^n$ is continuous. This inverse map is defined by

$$\phi_f(s) = \operatorname{argmax}_{x \in \mathbb{R}^n} c_s(x).$$

Given $s = \Phi_f(x)$, let $(s_n)_n$, $s_n \in \mathcal{S}_f$, be a sequence converging to s in L^2 -norm, such that $\lim_{n \rightarrow \infty} \phi_f(s_n) = y$. Then, $\tau_x f = \tau_y f$, and thus $\tau_{x-y} f = f$ and, since $f \neq 0$, we get that $x = y$ by Lemma 1(i), i.e.,

$$\lim_{n \rightarrow \infty} \phi_f(s_n) = \phi_f(s)$$

and this concludes the proof of Theorem 2.

ACKNOWLEDGMENTS

Pierre Kornprobst (INRIA, NeuroMathComp) is sincerely acknowledged for an appraisal of this work and for pointing out a few omissions in the math. Sincere thanks also go to the three reviewers for their critical appraisals, kind suggestions, and for the fact that they pointed out to the authors a disturbing unclarity in the original manuscript. Both authors were supported by the Research Grant Program of the Faculty of Science, University of Copenhagen (Research Grant 10-05087). Marco Loog received additional support from the Innovational Research Incentives Scheme of the Netherlands Research Organization (NWO, VENI-Grant 639.021.611).

REFERENCES

- [1] T. Lindeberg, "Feature Detection with Automatic Scale Selection," *Int'l J. Computer Vision*, vol. 30, no. 2, pp. 79-116, 1998.
- [2] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of Interest Point Detectors," *Int'l J. Computer Vision*, vol. 37, no. 2, pp. 151-172, 2000.
- [3] N. Sebe and M. Lew, "Comparing Salient Point Detectors," *Pattern Recognition Letters*, vol. 24, nos. 1-3, pp. 89-96, 2003.
- [4] W. Förstner and E. Gülch, "A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Circular Features," *Proc. Int'l Soc. for Photogrammetry and Remote Sensing Intercommission Conf. Fast Processing of Photogrammetric Data*, pp. 281-305, 1987.
- [5] C. Harris and M. Stephens, "A Combined Corner and Edge Detector," *Proc. Fourth Alvey Vision Conf.*, pp. 147-151, 1988.
- [6] B. Triggs, "Detecting Keypoints with Stable Position, Orientation, and Scale under Illumination Changes," *Proc. Eighth European Conf. Computer Vision*, pp. 100-113, 2004.
- [7] L. Rosenthaler, F. Heitger, O. Kubler, and R. von der Heydt, "Detection of General Edges and Keypoints," *Proc. Second European Conf. Computer Vision*, pp. 78-86, 1992.
- [8] D. Lisin, E. Riseman, and A. Hanson, "Extracting Salient Image Features for Reliable Matching Using Outlier Detection Techniques," *Proc. Third Int'l Conf. Computer Vision Systems*, 2003.
- [9] K. Walker, T. Cootes, and C. Taylor, "Locating Salient Object Features," *Proc. British Machine Vision Conf.*, pp. 557-566, 1998.
- [10] J. Shi and C. Tomasi, "Good Features to Track," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 593-600, 1994.
- [11] B. ter Haar Romeny, *Front-End Vision and Multi-Scale Image Analysis*. Kluwer Academic Publishing, 2003.
- [12] J. Noble, "Finding Corners," *Image and Vision Computing*, vol. 6, no. 2, pp. 121-128, 1988.
- [13] K. Rohr, "Modelling and Identification of Characteristic Intensity Variations," *Image and Vision Computing*, vol. 10, no. 2, pp. 66-76, 1992.
- [14] K. Mikolajczyk and C. Schmid, "Scale & Affine Invariant Interest Point Detectors," *Int'l J. Computer Vision*, vol. 60, no. 1, pp. 63-86, 2004.
- [15] T. Brox, J. Weickert, B. Burgeth, and P. Mrázek, "Nonlinear Structure Tensors," *Image and Vision Computing*, vol. 24, no. 1, pp. 41-55, 2006.
- [16] I. Laptev, "On Space-Time Interest Points," *Int'l J. Computer Vision*, vol. 64, no. 2, pp. 107-123, 2005.
- [17] P. Montesinos, V. Gouet, and R. Deriche, "Differential Invariants for Color Images," *Proc. 14th Int'l Conf. Pattern Recognition*, pp. 838-840, 1998.
- [18] N. Bruce, "Features that Draw Visual Attention: An Information Theoretic Perspective," *Neurocomputing*, vol. 65, pp. 125-133, 2005.
- [19] J. Fecteau and D. Munoz, "Saliency, Relevance, and Firing: A Priority Map for Target Selection," *Trends in Cognitive Sciences*, vol. 10, no. 8, pp. 382-390, 2006.
- [20] D. Gao, V. Mahadevan, and N. Vasconcelos, "On the Plausibility of the Discriminant Center-Surround Hypothesis for Visual Saliency," *J. Vision*, vol. 8, no. 7, pp. 1-18, 2008.
- [21] L. Itti and C. Koch, "Computational Modeling of Visual Attention," *Nature Rev. Neuroscience*, vol. 2, no. 3, pp. 194-203, 2001.
- [22] Z. Lingyun, M. Tong, and G. Cottrell, "Information Attracts Attention: A Probabilistic Account of the Cross-Race Advantage in Visual Search," *Proc. 29th Ann. Cognitive Science Conf.*, 2007.
- [23] L. Griffin, "The Second Order Local Image-Structure Solid," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 8, pp. 1355-1366, Aug. 2007.
- [24] J. Koenderink and A. van Doorn, "Generic Neighborhood Operators," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 6, pp. 597-605, June 1992.
- [25] T. Kadir and M. Brady, "Saliency, Scale and Image Description," *Int'l J. Computer Vision*, vol. 45, no. 2, pp. 83-105, 2001.

- [26] M. Loog, "On an Elementary Definition of Visual Saliency [Abstract]," *Perception*, vol. 37, no. suppl., p. 4, 2008.
- [27] L. Itti and P. Baldi, "Bayesian Surprise Attracts Human Attention," *Vision Research*, vol. 49, pp. 1295-1306, 2009.
- [28] R. Rosenholtz, "A Simple Saliency Model Predicts a Number of Motion Popout Phenomena," *Vision Research*, vol. 39, no. 19, pp. 3157-3163, 1999.
- [29] A. Torralba, "Modeling Global Scene Factors in Attention," *J. Optical Soc. Am. A*, vol. 20, no. 7, pp. 1407-1418, 2003.
- [30] G. Heidemann, "Focus-of-Attention from Local Color Symmetries," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 7, pp. 817-830, July 2004.
- [31] L. Itti et al., "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254-1259, Nov. 1998.
- [32] O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau, "A Coherent Computational Approach to Model Bottom-up Visual Attention," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 802-817, May 2006.
- [33] L. Matthies, M. Maimone, A. Johnson, Y. Cheng, R. Willson, C. Villalpando, S. Goldberg, A. Huertas, A. Stein, and A. Angelova, "Computer Vision on Mars," *Int'l J. Computer Vision*, vol. 75, no. 1, pp. 67-92, 2007.
- [34] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, second ed. Springer, 2006.
- [35] L. Florack, *Image Structure*. Kluwer Academic Publishers, 1997.
- [36] W. Rudin, *Functional Analysis*. McGraw-Hill Int'l Ed., 1991.
- [37] M. Markou and S. Singh, "Novelty Detection: A Review-Part 1: Statistical Approaches," *Signal Processing*, vol. 83, no. 12, pp. 2481-2497, 2003.
- [38] T. Lindeberg and J. Gårding, "Shape-Adapted Smoothing in Estimation of 3D Shape Cues from Affine Deformations of Local 2D Brightness Structure," *Image and Vision Computing*, vol. 15, no. 6, pp. 415-434, 1997.
- [39] K. Rohr, "On 3D Differential Operators for Detecting Point Landmarks," *Image and Vision Computing*, vol. 15, no. 3, pp. 219-233, 1997.
- [40] V. Navalpakkam and L. Itti, "Modeling the Influence of Task on Attention," *Vision Research*, vol. 45, no. 2, pp. 205-231, 2005.
- [41] L. Zhang, M. Tong, T. Marks, H. Shan, and G. Cottrell, "SUN: A Bayesian Framework for Saliency Using Natural Statistics," *J. Vision*, vol. 8, no. 7, pp. 1-20, 2008.
- [42] E. Knudsen, "Fundamental Components of Attention," *Ann. Rev. Neuroscience*, vol. 30, pp. 57-78, 2007.
- [43] W. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry*. Academic Press, 1975.
- [44] S. Gallot, D. Hulin, and J. Lafontaine, *Riemannian Geometry*. Springer-Verlag, 1990.
- [45] R. Adams and J. Fournier, *Sobolev Spaces*. Academic Press, 1975.

► For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.