Scale invariance for evolved interest operators

Leonardo Trujillo and Gustavo Olague

Proyecto Evovisión,
Centro de Investigación Científica y de Educación Superior de Ensenada, Km. 107
Carretera Tijuana-Ensenada, 22860, Ensenada, BC, México
trujillo@cicese.mx,olague@cicese.mx
home page: http://ciencomp.cicese.mx/evovision

Abstract. This work presents scale invariant region detectors that apply evolved operators to extract an interest measure. We evaluate operators using their repeatability rate, and have experimentally identified a plateau of local optima within a space of possible interest operators Ω. The space Ω contains operators constructed with Gaussian derivatives and standard arithmetic operations. From this set of local extrema, we have chosen two operators, obtained by searching within Ω using Genetic Programming, that are optimized for high repeatability and global separability when imaging conditions are modified by a known transformation. Then, by embedding the operators into the linear scale space generated with a Gaussian kernel we can characterize scale invariant features by detecting extrema within the scale space response of each operator. Our scale invariant region detectors exhibit a high performance when compared with state-of-the-art techniques on standard tests.

1 Background

Current trends in Computer Vision (CV) are to adopt a simplified approach to address the problems of: object detection/recognition, content based image retrieval and image indexing [10]. This approach works with image information extracted directly from local image features which makes it robust to partial object occlusions and eliminates the need for prior segmentation. However, the approach does require the detection of stable image features that correspond to *visually interesting regions*, of which interest points are the most widely known [11–13]. Interest regions exhibit a high level of variation with respect to a local measure that is extracted using a particular image operator. Hence, different region detectors define different operators that extract an interest measure for every image pixel. After applying the interest operator local extrema are selected as interest regions. The main characteristic, and the only one for which a reliable performance metric exists, that is expected from an interest region operator is stability under changes in viewing conditions quantified by the repeatability rate [11]. Stability is evaluated under different kinds of transformations, which include: *translation, illumination change, rotation, scale change and projective transformations*. Interest region detectors invariant to the first three types of image transformations are better know as interest point detectors [12, 13], invariance to the first four are scale invariant region detectors [9], while invariance to all are know as affine covariant region detectors [9]. Previous work by Trujillo and Olague [12, 13] proposed a novel approach to construct optimized interest point detectors using Genetic Programming (GP) as an optimization engine and the repeatability rate as part of the fitness. The present work extends that contribution by embedding evolved operators into a linear scale space to detect scale invariant regions [6], and presents the following contributions to the field of automatic feature extraction with Evolutionary Computation. First, this work characterizes a conceptual search space for interest operators, that are applicable to different types of CV applications. Second, we identify how artificial evolution automatically rediscovered basic image analysis techniques that have long been considered as possible models for low level vision in biological visual systems. Finally, scale invariant region detectors are presented that extract an interest measure based on evolved operators with better performance than manmade designs, and a simpler structure.

Interest Operators. These are functions that operate on a local neighborhood of every image pixel and extract a corresponding interest measure K. This operation produces an *interest image* which is subsequently thresholded to detect local extrema. Popular interest operators, designed to detect interest points, include [5, 4]:

$$K_{Harris\&Stephens}(\mathbf{x}) = det(A) - Tr(A)^2 \ ,$$

$$K_{Forstner}(\mathbf{x}) = \frac{det(A)}{Tr(A)} \ ,$$

where A is the local autocorrelation matrix [9] given by

$$A(\mathbf{x}, \sigma_I, \sigma_D) = \sigma_D^2 \cdot G_{\sigma_I} * \begin{bmatrix} L_x^2(\mathbf{x}, \sigma_D) & L_x L_y(\mathbf{x}, \sigma_D) \\ L_x L_y(\mathbf{x}, \sigma_D) & L_y^2(\mathbf{x}, \sigma_D) \end{bmatrix} \ ,$$

where σ_I and σ_D are the derivation and integration scales respectively, L_u is the Gaussian derivative in direction u and G_σ is a Gaussian smoothing function with standard deviation σ. Other interest measures are related to the curvature at each point, such as the determinant of the Hessian proposed by Beaudet [1]:

$$K_{Beaudet}(\mathbf{x}) = I_{xx}(\mathbf{x}, \sigma_D) \cdot I_{yy}(\mathbf{x}, \sigma_D) - I_{xy}^2(\mathbf{x}, \sigma_D) \ .$$

Wang and Brady [14] propose an interest measure related to the curvature of an edge using the Laplacian along with the gradient magnitude,

$$K_{Wang\&Brady}(\mathbf{x}) = (\nabla^2 I)^2 - s|\nabla I|^2 \ .$$

Constructing Interest Operators with Genetic Programming. Early contributions to this problem include [2, 3], however those works do not define a proper fitness measure and their results are neither reusable or general. A novel framework to automatically synthesize interest operators with GP, that overcomes the shortcomings in [2,3], is presented in [12,13]. From a careful analysis of the above mentioned operators, as well as others, the authors define the following Function and Terminal sets that would allow us to construct any of them, as well as a vast amount of unknown operators.

$$F = \left\{ +, -, |-|, *, /, I^2, \sqrt{I}, log_2, \frac{I}{2}, G_{\sigma=1}, G_{\sigma=2} \right\} \ , \tag{1}$$

$$T = \{I, L_x, L_{xx}, L_{xy}, L_{yy}, L_y\} \ , \tag{2}$$

where F and T are the function and terminal set respectively. Some authors [4,14,1] do not use Gaussian derivatives, however F is defined in this way because they are less susceptible to noise. Furthermore, an appropriate evaluation function $f(o)$ should depend on each operator's repeatability rate $r_{o,J}(\epsilon)$ for an operator o, on an image sequence J. Therefore, the fitness landscape is

$$f(o) \propto r_{o,J}(\epsilon) \ , \tag{3}$$

where ϵ is an error threshold, see [11–13].

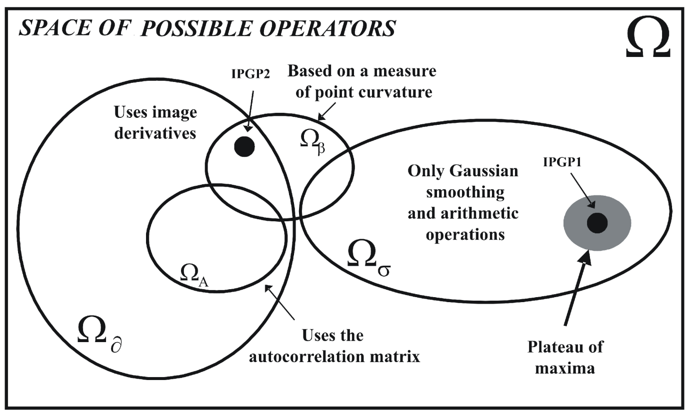

Fig. 1. Space of possible interest operators.

Figure 1 represents a high level view of the space Ω of possible interest operators constructed with the above mentioned primitives. A subspace $\Omega_\delta \subset \Omega$ represents the space of possible operators that use image derivatives explicitly, taken from T, to obtain their interest measure; $\Omega_\sigma \subset \Omega$ only contains operators that use Gaussian smoothing and arithmetic operations included in F. The subspaces Ω_δ and Ω_σ group operators based on their their genotype and not their phenotype. Figure 1 also shows where we might find the subspace of operators that rely on measures pertaining to the local autocorrelation matrix Ω_A, or that extract a measure related to surface curvature Ω_β [1], along with the two opera-

[1] Ω_β contains operators with similar functionality and its intersection with other subspaces, those based on structure, may or may not be empty.

tors presented in [12, 13], $IPGP1$ and $IPGP2$ [2]. These operators outperformed or matched all previous man made designs on standard tests [16],

$$K_{IPGP1}(\mathbf{x}) = G_{\sigma=2} * (G_{\sigma=1} * I - I) \ , \quad (4)$$

$$K_{IPGP2}(\mathbf{x}) = G_{\sigma=1} * (L_{xx}(\mathbf{x}, \sigma_D = 1) \cdot L_{yy}(\mathbf{x}, \sigma_D = 1) - L^2_{xy}(\mathbf{x}, \sigma_D = 1)) \ . \quad (5)$$

$IPGP1$ identifies corners, edges and blobs with salient low intensity regions. Its additive inverse extracts salient high intensity regions. $IPGP2$ on the other hand is a modified version of the operator proposed by Beaudet [1], similar to the improvements made in [11] to the Harris and Stephens detector, what the authors called *Improved Harris*. Further experimental runs of the GP search have identified a plateau of local maxima in the neighborhood of $IPGP1$. Here, we present a close neighbor, both in the function space and fitness space, of $IPGP1$ that we name $IPGP1^*$,

$$K_{IPGP1^*}(\mathbf{x}) = G_{\sigma=2} * |G_{\sigma=1} * I - I| \ . \quad (6)$$

$IPGP1^*$ identifies maxima related to both $IPGP1$ and its additive inverse.

Proposition 1. *Both $IPGP1$ and $IPGP1^*$ are proportional to DoG (Difference-off-Gaussian) filters, if we assume that image I is derived from an unknown image $\hat{I}$ blurred with a Gaussian of unknown standard deviation $\hat{\sigma}$ such that $I = G_{\hat{\sigma}} * \hat{I}$, and*

$$G_\sigma * I - I = G_\sigma * G_{\hat{\sigma}} * \hat{I} - G_{\hat{\sigma}} * \hat{I} \propto G_{\sigma+\hat{\sigma}} * \hat{I} - G_{\hat{\sigma}} * \hat{I} = (G_{\sigma+\hat{\sigma}} - G_{\hat{\sigma}}) * \hat{I} \ . \quad (7)$$

Therefore, $IPGP1$ and $IPGP1^*$ are approximations of the 2D LoG function.

2 Scale Space Analysis

One of the basic problems in CV is to determine the scale at which image information should be analyzed. Different real world structures are only appreciable and relevant at certain scales and lack importance at others. Thus, a solution to this problem has been proposed by applying the concept of scale-space, which allows us to work explicitly with the scale selection problem while also simplifying image analysis by only focusing on *interesting* scales. For a useful, if not rigorous, concept of scale we turn to one of the most important contributions in scale-space theory by Lindeberg [6].

> *The scale parameter should be interpreted only as an abstract scale parameter implying a weak ordering property of objects of different size without any direct mapping from its actual value to the size of features in a signal represented at that scale.*

A multi-scale representation of an image is obtained by embedding it within a family of derived signals which depend on the lone scale parameter t [6].

[2] $IPGP$ is an acronym for *Interest Point Detection with GP*.

Definition 1. *Given a signal $f : \Re^D \rightarrow \Re$, the linear scale-space representation $L : \Re^D \times \Re \rightarrow \Re$ of f is given by the solution to the diffusion equation*

$$\delta_t L = \frac{1}{2}\nabla^2 L = \frac{1}{2}\sum_{i=1}^{D} \delta_{x_i x_i} L \ , \tag{8}$$

with the initial condition $L(\cdot;0) = f(\cdot)$, for which the Green function solution is the Gaussian kernel. Equivalently, it is possible to define the scale-space as the family of derived signals obtained by convolving a signal f with Gaussian filters at different scales t (standard deviation),

$$L(\cdot;t) = G_t * f(\cdot) \ . \tag{9}$$

Lindeberg notes that the scale-space-representation could be taken as a canonical model for biological vision due to results in neurophysiological studies [15]. Now, to determine the scale at which image features should be analyzed Lindeberg presents a **Principle for scale selection** [7]: *"In the absence of other evidence, assume that a scale level, at which some (possibly non-linear) combination of normalized derivatives assumes a local maximum over scales, can be treated as reflecting a characteristic length of a corresponding structure in the data"*. Normalized derivatives are invariant at different scales [6]. In practice however, Lindeberg concludes that the usefulness of the principal for scale selection *"... must be verified empirically, and with respect to the type of problem it is to be applied to"*. Hence, we can expect that an experimental GP search for candidate interest operators is a valid approach to construct a scale invariant detector based on a *"possibly non-linear combination of normalized derivatives"*. Furthermore, it is possible to contemplate that the GP search will be biased to simplified approximate measures, such as the approximation of the LoG by way of DoG filters. This can be induced in a GP search by applying specific genetic or selection operators that help keep evolved operators simple [12, 13].

Characteristic Scale From an algorithmic point of view, selecting a characteristic scale for local image features is a process in which local extrema of a function response are found over different scales [7]. Given a function, or interest operator, $F(\mathbf{x}, t_i)$ that computes an interest measure for each image pixel $\mathbf{x}$ at different scales t_i, we can assume that the characteristic scale at $\mathbf{x}$ is t_n if

$$\begin{aligned} F(\mathbf{x}, t_n) > \sup\left\{F(\mathbf{x_W}, t_n), F(\mathbf{x_W}, t_{n-1}) F(\mathbf{x_W}, t_{n+1}) | \forall \mathbf{x_W} \in \mathbf{W}, \mathbf{x_W} \neq \mathbf{x}\right\} \\ \wedge F(\mathbf{x}, t_n) > h \ , \end{aligned} \tag{10}$$

where h is a threshold, and $\mathbf{W}$ is a $n \times n$ neighborhood around $\mathbf{x}$. This process, similar to what is done for interest point detection, will return a set of local scale invariant regions, each centered on an image pixel $\mathbf{x}$.

3 Scale Invariant Detectors

Now that we have defined the concept of characteristic scale and a basic methodology on how to obtain it, we can move on to present our proposed scale invariant detectors. Here, we are interested in detectors derived from the scale-space representation. As a starting point, we turn to Mikolajczyk and Schmid [9] who gave a comparison of different scale invariant detectors, including: DoG [8], Hessian, Laplacian [7] and Harris-Laplace [9]. From this comparison the authors experimentally concluded that, as expected, the DoG and Laplacian gave very similar results and that the Harris-Laplace detector gave the highest repeatability rate for scale change transformations. As mentioned before, we will present detectors based on the $IPGP1$ and $IPGP1^*$ interest operators, which according to proposition 1 are proportional to DoG filters. However, we present a different algorithmic implementation that maintains the basic structure of the operator and produces better performance. Scale invariant detection using DoG as proposed by Lowe [8] uses a scale space pyramid and DoG filters are applied between adjacent scales. Here, our $IPGP$ based detectors will perform DoG filtering between each scale and the base image, with scale $t = 0$, contrasting with the implementation in [8] in which both Gaussian functions of the DoG filter are modified sequentially. In order to apply our evolved operators within a scale-space analysis we must modify their definition by including the scale parameter, such that

$$K_{IPGP1_t}(\mathbf{x};t) = G_{t_i} * (G_{t_i} * I - I) \ , \tag{11}$$

$$K_{IPGP1^*{}_t}(\mathbf{x};t) = G_{t_i} * (G_{t_i} * I - I) \ , \tag{12}$$

where t_i is the current scale, $i = 1...N$ with N the total number of scales analyzed. Now our operators are scale dependent and will return a different interest measure for each pixel at different scales. Hence, it is now possible to apply the characteristic scale selection criteria. Our operators avoided the need for normalized derivatives and are more efficient than other detectors as reported in [9]. Note that we are not using operators evolved explicitly for high repeatability under scale change. However, current state-of-the-art detectors rely on interest point operators embedded into the linear scale-space, the same approach we are taking. This can be seen as having a relationship with the area of interactive evolution where user input guides the selection process, as we did by selecting which operators to employ.

Implementation. Because the process is straightforward, the only requirements are to establish a set of parameters that are defined empirically, as is the case for all region detectors. We set $N = 20$ and $t_i = 1.2^i$; the size of our scale neighborhoods $\mathbf{W}$ was set to $n = 5$, and our thresholds h were chosen experimentally. For comparison purposes, we use the Harris-Laplace and Hessian-Laplace detectors, using the authors binaries downloaded from the Vision Geometry Group website [16], along with five image sequences: Laptop, BIP, VanGogh, Asterix and Boat; where the first four are sequences that only present scale change transformations while the fifth has both scale and rotation.

Results. Figure 2 is a qualitative comparison that shows interest regions extracted by each detector. It is possible to observe how $IPGP1$ and its additive inverse extract complementary regions, while $IPGP1^*$ extracts a combination of maxima from both. Furthermore, the *IPGP* operators and the Harris and Hessian based methods exhibit similarities. Figure 3 is a quantitative comparison of each detector, it presents the repeatability rate on 5 different image sequences. The performance graphics plot the images in the sequence and the repeatability rate with respect to the base image. Each image in the sequence is progressively transformed, i.e., the view point of image 4 is closer to that of the base image than the viewpoint of image 10 in the BIP sequence [16]. All detectors exhibit similar performance patterns. However, we can appreciate that the detectors based on evolved operators are slightly better on most sequences.

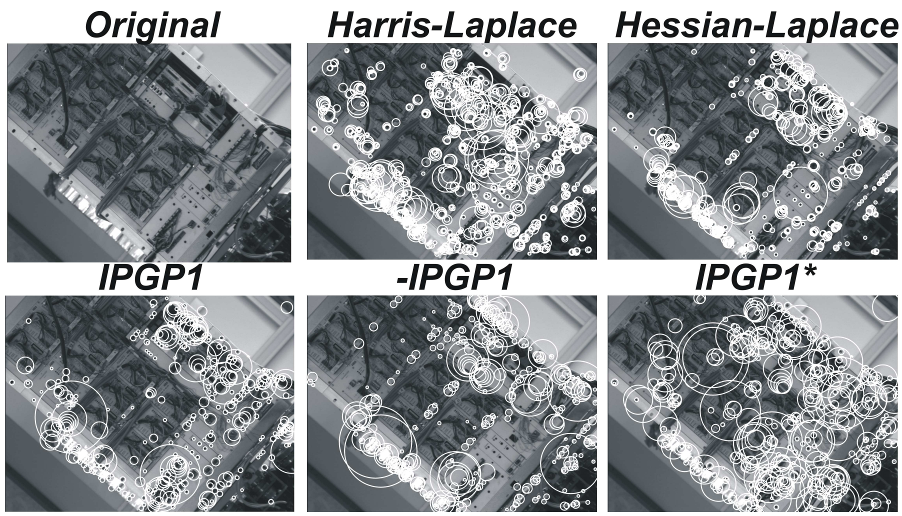

Fig. 2. Sample image regions. Regions are shown with circles of radius $r = 3 \cdot t_n$ pixels, with t_n the regions characteristic scale.

4 Discussion and Future Work

This paper presented scale invariant detectors based on operators optimized for high repeatability using GP. The detectors were embedded into a linear scale space generated with a Gaussian kernel and compared with state-of-the-art detectors. Results show that our detectors are, on average, better than other detectors based on their repeatability rate, while at the same time maintaining a simpler structure. Our results show that simple operators found by simulated evolution can outperform more elaborate manmade designs. This interesting result substantiates the belief that evolution will always find the simplest and most apt solution to a given problem. This is made possible by correctly framing the evolutionary search process with a fitness function that promotes the extraction of highly repeatable regions, a property that is useful in many vision applications. As possible future work, there are two main extensions that should be

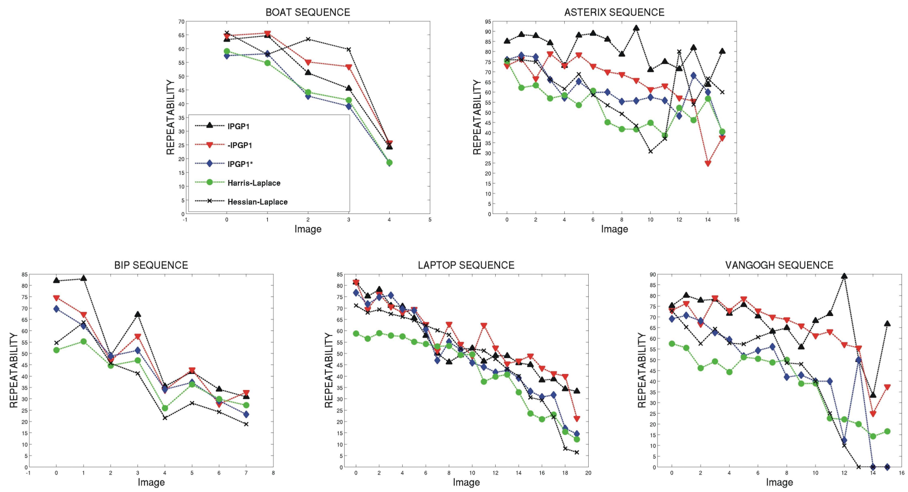

Fig. 3. Repeataility for each detector in our comparison for each image sequence.

explored. First, employ an evolutionary search process that directly takes into account scale space analysis in its fitness evaluation. Second, extend the use of evolved operators to extract affine covariant features, a much more challenging problem in CV.

Acknowledgments: Research funded by UC MEXUS-CONACyT Collaborative Research Grant 2005 through the project "Intelligent Robots for the Exploration of Dynamic Environments". First author supported by scholarship 174785 from CONACyT. This research was also supported by the LAFMI project.

References

1. P. R. Beaudet: Rotational invariant image operators, in Proceedings of the 4th International Joint Conference on Pattern Recognition, (1978) 579–583.
2. M. Ebner: On the Evolution of Interest Operators using Genetic Programming, in Late Breaking Papers at EuroGP'98, (1998) 6–10.
3. M. Ebner and A. Zell: Evolving Task Specific Image Operator, in Proceedings from the First European Workshops, EvoIASP'99 and EuroEcTel'99, Lecture Notes in Computer Science Vol. 1596, (Springer-Verlag 1999) 74–89.
4. W. Förstner: A framework for low level feature extraction, in Proceedings of the 3rd European Conference on Computer Vision, (1994) 383–394.
5. C. Harris and M. Stephens: A combined corner and edge detector, in Alvey Vision Conference, (1988) 147–151.
6. T. Lindeberg: Discrete Scale-Space Theory and the Scale-Space Primal Sketch. PhD thesis, Computational Vision and Active Perception Laboratory (CVAP) Royal Institute of Technology, Sweden, (1991).
7. T. Lindeberg: Feature detection with automatic scale selection. Int. J. Comput. Vision, **30**(2) (1998) 79–116.
8. D.G. Lowe: Object recognition from local scale-invariant features, in Proceedings of CVPR 1999, Vol. 2, (IEEE Computer Society 1999) 1150–1157.
9. K. Mikolajczyk and C. Schmid: Scale and Affine Invariant Interest Point Detectors. Int. J. Comput. Vision, **60**(1) (2004) 63–86.
10. K. Mikolajczyk and C. Schmid: A Performance Evaluation of Local Descriptors. IEEE Trans. Pattern Anal. Mach. Intell., **27**(10) (2005) 1615-1630.
11. C. Schmid, R. Mohr, and C. Bauckhage: Evaluation of interest point detectors. Int. J. Comput. Vision, **37**(2) (2000) 151–172.
12. L. Trujillo and G. Olague: Synthesis of interest point detectors through genetic programming, in Proceedings from GECCO 2006, Vol.1, (ACM Press 2006), 887–894.
13. L. Trujillo and G. Olague: Using Evolution to Learn How to Perform Interest Point Detection, in Proceedings from ICPR 2006, Vol. 1, (IEEE Computer Society 2006), 211–214.
14. H. Wang and J.M. Brady: Corner detection for 3d vision using array processors, in Proceedings from BARNAIMAGE-91, (Springer-Verlag 1991).
15. R.A. Young: Simulation of the Human Retinal Function with the Gaussian Derivative Model, in Proceedings of the 1986 IEEE Conference on Computer Vision and Pattern Recognition, (1986) 564–569.
16. Visual Geometry Group: http://www.robots.ox.ac.uk/ vgg/research/