
Automated design of image operators that detect interest points

Leonardo Trujillo

trujillo@cicese.mx

Proyecto Evovisión, Departamento de Ciencias de la Computación, División de Física Aplicada, Centro de Investigación Científica y de Educación Superior de Ensenada, Km. 107 Carretera Tijuana-Ensenada, 22860, Ensenada, BC, México.

Gustavo Olague

olague@cicese.mx

Proyecto Evovisión, Departamento de Ciencias de la Computación, División de Física Aplicada, Centro de Investigación Científica y de Educación Superior de Ensenada, Km. 107 Carretera Tijuana-Ensenada, 22860, Ensenada, BC, México.

Abstract

This work describes how evolutionary computation can be used to synthesize low-level image operators that detect interesting points on digital images. Interest point detection is an essential part of many modern computer vision systems that solve tasks such as object recognition, stereo correspondence and image indexing, to name but a few. The design of the specialized operators is posed as an optimization/search problem that is solved with genetic programming (GP), a strategy still mostly unexplored by the computer vision community. The proposed approach automatically synthesizes operators that are competitive with state-of-the-art designs, taking into account an operator's geometric stability and the global separability of detected points during fitness evaluation. The GP search space is defined using simple primitive operations that are commonly found in point detectors proposed by the vision community. The experiments described in this paper extend previous results (Trujillo and Olague, 2006a,b) by presenting 15 new operators that were synthesized through the GP-based search. Some of the synthesized operators can be regarded as improved man-made designs because they employ well-known image processing techniques and achieve highly competitive performance. On the other hand, the GP search also generates what can be considered as unconventional operators for point detection, these results provide a new perspective to feature extraction research.

Keywords

Feature detection, genetic programming, interest points, computer vision.

1 Introduction

Historically, vision systems would include an image segmentation stage during low-level processing. Segmentation attempts to group image pixels into fully connected and homogeneous regions. Each region, it is expected, will cover actual objects or clearly identifiable concepts, e.g., *car* or *face*. However, segmentation is very difficult in the general case because it is an ill posed problem. Hence, a recent approach used by computer vision (CV) researchers is to adopt a simplified methodology for problem domains such as object detection/recognition, content based image retrieval, and image indexing. Essentially, the segmentation stage is eliminated by only focusing on local and relatively small amounts of image information. The approach that is being made

reference to was introduced by Schmid and Mohr (1997) and subsequently extended by the work of Lowe (1999) and others. This paper presents a novel perspective on how to solve the problem of identifying locally salient information through an evolutionary process. Thus, in order for the reader to grasp the relevance of the present contribution, a brief overview on how this process is normally studied follows.

The approach described by (Schmid and Mohr, 1997; Lowe, 1999) proceeds as follows. First, *interesting* or salient image regions must be detected. Afterwards, discriminative descriptors are used to characterize the local neighborhood around each detected feature, such as the popular SIFT descriptor (Lowe, 1999), a PCA method (Ke and Sukthankar, 2004), or an image description based on local signal regularity (Trujillo et al., 2007b). Finally, detected features and their respective descriptors are used to construct generative (Trujillo et al., 2007a) or discriminant (Lowe, 1999) scene and object models. When the previously described scheme is employed it is paramount to use *feature detectors* that can identify stable and informative image regions. Detectors that fulfill these requirements are known as interest region detectors, of which interest point detectors are the most well-known and widely used by vision systems.

Nevertheless, the overwhelming majority of interest point detectors are the product of a detailed human analysis which usually produces a single design. Conversely, evolutionary computation (EC) provides a framework that allows for the automated design of solutions in different problem domains (Poli et al., 2008; Cagnoni et al., 2007). Hence, multiple solutions can be obtained by simply running the evolutionary process several times. This can be achieved by correctly framing the problem statement through two essential aspects: the fitness function and the search space. In the literature few examples exist where feature detectors have been automatically synthesized (Ebner, 1998; Ebner and Zell, 1999). However, in those examples the search problem was not framed correctly and have thus failed to produce truly competitive or general results. It was not until previous work by Trujillo and Olague (2006a,b) that the automatic synthesis of interest point operators was framed in such a way as to evolve truly human-competitive solutions.

1.1 Research contributions

This paper outlines the following research contributions:

1. First, the automated design of an interest point detector is explained, where the problem is stated in terms of an optimization/search process using GP. The explanation presented extends the one of (Trujillo and Olague, 2006a,b).
2. Second, a new set of high-performance interest point detectors is presented; a total of 15 new and sometimes unprecedented designs. Each detector can be used in situations where vision-systems require interest point detection. In fact, the GP search produced several operators that are very simple. This is in contrast to common published GP results where the parsimony principle is not fulfilled. Here, the operators designed by artificial evolution are able to outperform more elaborate man-made designs. This was unexpected given the unquestionable difficulty of the detection problem. In this way, it can be argued that simulated evolution tends to find the simplest and most apt solution when appropriate fitness criteria are defined.
3. Third, a detailed analysis of the search space was undertaken, based on a large amount of experimental work. During the analysis, special attention was given to

the convergence tendency of the GP-search towards optima centered around the Laplacian-of-Gaussian (LoG) and Difference-of-Gaussian (DoG) filters.

4. Finally, an analysis of each successful experiment is presented. The structure of each of the best operators is described along with their respective performance, and insights regarding the dynamics of each evolving population are given.

The remainder of this paper proceeds as follows: Section 2 gives a brief overview of related research to appropriately frame the contributions of the present work in terms of what is now known as evolutionary computer vision (Olague et al., 2006a; Cagnoni et al., 2007). Section 3 defines the concept of *interest points*, it also reviews how interest point detection is performed, and describes how this process can be evaluated. The proposed GP-based methodology is detailed in Section 4. Afterwards, experimental results are presented in Section 5 with a detailed discussion. Finally, the last section contains a summary and concluding remarks.

2 Related work

The capability for visual perception, rediscovered many times by natural evolution, is still a daunting problem for modern CV. Vision requires the fast processing of multidimensional and redundant information, it involves low-level data acquisition and high-level cognition. This presents researchers with many complex problems that are still unsolved, prompting some to employ *black-box* computation methods. In particular, EC has begun to produce promising results because of its intrinsic ability to search within high dimensional and multimodal spaces. This relatively new perspective has produced a promising area of research that is reformulating well-known vision problems and proposing highly creative solutions (Olague et al., 2006a; Cagnoni et al., 2007). Published literature in this area includes work on: image segmentation (Poli, 1996), camera network design (Olague and Mohr, 2002), dense stereo correspondence (Olague et al., 2006b), 3D reconstruction (Olague and Puente, 2006), and feature selection (Sun et al., 2004; Hernández et al., 2007; Trujillo et al., 2007a), to name but a few. However, the domain that is directly related to the present work is that pertaining to feature synthesis, where EC algorithms have been used to identify specific types of image features. For example, Lin and Bhanu (2005) perform object detection in SAR images using GP in a cooperative co-evolutionary framework. They start with simple features that are useful for the problem domain, what they call *primitive features*, as a terminal set. After, novel features are synthesized using this a priori information. The use of domain related information accelerates the evolutionary search and confines it to promising regions within the search space. Zhang et al. (2003) use GP to perform multi-class detection of small objects present in large images, with domain independent pixel statistics as the terminal set. Both of the above mentioned contributions do not address low-level problems, they focus on providing novel methodologies for higher level applications, and each time a new problem instance is presented a new run of the algorithm is required. Other works attempt to synthesize algorithmic solutions, using a single evolutionary run, that can later be used by different vision systems. For instance, Ebner (1998) focuses on constructing an equivalent to the Moravec interest point detector (Moravec, 1977) using GP. Experimental results shows a 15% localization error between interest point detection with the evolved operator and that obtained using the Moravec detector. In another work, Ebner and Zell (1999) present an evolved operator optimized for optical flow estimation. The main drawbacks of these contributions is that they fail to

define appropriate fitness criteria, and produce specialized functions that are difficult to analyze and interpret.

The current paper describes a methodology that synthesizes image operators that detect low-level features, also called *interest points*. The approach evolves operators that can be used by vision applications that need to detect salient pixels that are stable and sparsely distributed within the image. The experimental comparisons presented here show highly competitive, and in most cases superior, performance compared to man-made designs. Hence, it is reasonable to assume that a wide variety of vision systems can benefit from the automated design process described in this work.

3 Visually interesting points

Previously, an approach for vision-based recognition was described where the detection of stable image features is required (Schmid and Mohr, 1997; Lowe, 1999). Said features are relatively small compared to the whole image, they are usually called interest points because they convey *visually interesting information*. A measure of how interesting an image pixel is, can be extracted using a mapping $K : \mathbb{R}^+ \rightarrow \mathbb{R}$ or interest operator¹; where, different detectors employ different K ². Applying K to an image I produces what can be called an *interest image* I^* , see Figure 1. Afterwards, most detectors follow the same basic process: non-maxima suppression that eliminates pixels that are not local maxima, and a final thresholding that obtains the final set of features.

Schmid et al. (2000) proposed a performance measure that can be used to evaluate and compare different detectors known as the repeatability rate. This measure characterizes the geometric stability of a detector when the conditions produced by image acquisition change. Stable and easily repeatable points make them, in the words of Shi and Tomasi (1994), *good features to track*. In this context, visually interesting regions contain image pixels that exhibit a distinctive property that make them suitable for applications where specific scene points need to be tracked across multiple images, or identified under different viewing conditions. The stability of a detector has been evaluated under different kinds of image transformations: *illumination change, rotation, scale change* and *affine transformations*. Detectors invariant to the first two types of image transformations extract interest points (Trujillo and Olague, 2006a,b), invariance to the first three extract scale invariant regions (Trujillo and Olague, 2007), and detectors invariant to all identify affine covariant regions (Mikolajczyk et al., 2005).

3.1 Interest point detection

Interest point detection resulted from research devoted to corner detection, a taxonomy of which contains three basic classes: *contour methods* (Asada and Brady, 1986), *parametric model methods* (Olague and Hernández, 2005) and *image intensity methods* (Moravec, 1977; Beaudet, 1978; Kitchen and Rosenfeld, 1982; Harris and Stephens, 1988; Wang and Brady, 1991; Förstner and Gülich, 1987; Shi and Tomasi, 1994). The class of corner detectors that operate directly on the intensity image are more appropriately referred to as interest point detectors. While corners are point features located at line and surface junctions, such as L, T, Y and X junctions, interest points include these features as well as others that may lack a clear semantic interpretation. Relevant examples of interest point detectors are recalled next.

¹The domain of K is \mathbb{R}^+ because images do not contain negative intensity values.

²A *detector* refers to the complete algorithmic process that extracts interest points. On the other hand, an *operator* only computes the interest measure for each pixel.



Figure 1: A look at interest point detection: left, an input image I ; middle, *interest image* I^* ; right, detected points after non-maximum suppression and thresholding superimposed on I . Interest points were detected with the K_{IPGP2} operator.

For instance, some detectors base their interest measure K on the local autocorrelation matrix A , which characterizes the gradient distribution around each image pixel, with

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

where σ_D and σ_I are the derivation and integration scales respectively, $L_u(\mathbf{x}, \sigma_D)$ is the Gaussian derivative in direction u of image I at point \mathbf{x} given by

$$L_u(\mathbf{x}, \sigma_D) = \frac{\delta}{\delta u} G_{\sigma_D} * I(\mathbf{x}),$$

G_σ is a Gaussian smoothing function with standard deviation σ ; $\sigma_D = 1$ is used unless noted otherwise. Detectors based on A include those proposed by Förstner and Gülich (1987), Harris and Stephens (1988), and Shi and Tomasi (1994), with their corresponding interest measures given by

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

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

$$K_{Shi\&Tomasi}(\mathbf{x}) = \min \{\lambda_1, \lambda_2\},$$

where λ_1, λ_2 are the two eigenvalues of A ; the definition of A is taken from (Schmid et al., 2000), used with the *Improved Harris* detector described in that work. Beaudet (1978) proposed the determinant of the Hessian, which is proportional to the Gaussian curvature, as an interest measure,

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

where $I_u(\mathbf{x})$ is the image derivative in direction u . Wang and Brady (1991) characterize the curvature response using the Laplacian and the gradient magnitude,

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

Kitchen and Rosenfeld (1982) present an interest measure aimed at detecting image corners that is given by the product between the gradient magnitude and the magnitude of the change of direction of the gradient,

$$K_{Kitchen\&Rosenfeld}(\mathbf{x}) = \frac{I_{xx}(\mathbf{x})I_y^2(\mathbf{x}) + I_{yy}(\mathbf{x})I_x^2(\mathbf{x}) - 2I_{xy}(\mathbf{x})I_y(\mathbf{x})I_x(\mathbf{x})}{I_x^2(\mathbf{x}) + I_y^2(\mathbf{x})}.$$

All detectors follow the same process. First apply K to I and obtain the interest image I^* . Then, a pixel \mathbf{x} is tagged as an interest point if the following conditions hold,

$$K(\mathbf{x}) > \sup \{ K(\mathbf{x}_W) | \forall \mathbf{x}_W \in \mathbf{W}, \mathbf{x}_W \neq \mathbf{x} \} \wedge K(\mathbf{x}) > h , \quad (1)$$

where \mathbf{W} is a square neighborhood of size $n \times n$ around \mathbf{x} , and h is an empirically defined threshold. The first condition in Eq. (1) accounts for non-maximum suppression and the second is the thresholding step, the process is shown in Figure 1. Experiments in the current work use $n = 5$, while h is operator dependent. Several extensions of this process have been proposed. Laptev and Lindeberg (2003), for example, include a space-time analysis, and van de Weijer et al. (2006) incorporate color information. However, these improvements are not considered within the scope of this paper.

3.2 Evaluating interest point detectors

It is commonly assumed that interest points should fulfill the following properties: **global separability** between extracted points; **high information content** when compared to other pixels; and **stability**. The previous list is not exhaustive, nor is it meant to be rigorous. However, it does express desired properties that a detector is expected to fulfill when employed by a higher-level system. The *global separability* of extracted points suggests that on *most images* interest points will not be crowded together on isolated portions. This property will be image dependent and requires prior knowledge of the expected number of points and their position. Nevertheless, when interest points are used to build models of objects or images it is better to have a close to uniform sampling of the visual appearance of the imaged scene. A *high information content* refers to the uniqueness of the local neighborhood around each detected point. This property would facilitate interest point matching based on local descriptors. *Information content* is related with *global separability* because if points are cluttered together their description will be very similar and could thus be less unique. Finally, the *stability* of detected points is probably the most important criterion and the only one for which an accepted measure exists, the repeatability rate (Schmid et al., 2000).

3.2.1 Repeatability rate

The most common performance measure for interest point detectors is the repeatability rate that quantifies how detection is invariant to changes produced by image acquisition, such as rotations or scene illumination (Schmid et al., 2000). An interest point x_1 detected on image I_1 is repeated in image I_i if the corresponding point x_i is detected in image I_i ; for simplicity, both images are assumed to be of the same size. In the case of planar scenes, a relation between points x_1 and x_i can be established with the homography $H_{1,i}$, where

$$x_i = H_{1,i}x_1 . \quad (2)$$

The repeatability rate measures the number of repeated points between both images with respect to the total number of detected points. However, parts of image I_1 may not appear on the transformed image I_i . In order to account for this, repeated and detected points are only counted if they lie within the common parts of image I_1 and image I_i . Furthermore, a small amount of detection error needs to be taken into account because exact localization is not a necessity for most applications. A repeated point is said to be detected at pixel x_i if it lies within a given neighborhood of x_i of size ϵ , all our experiments use $\epsilon = 1.5$ pixels, see Figure 2. The set of point pairs (x_1^c, x_i^c) that lie in the common part of both images and correspond within an error ϵ is defined by

$$R_{I_i} (\epsilon) = \{(x_1^c, x_i^c) | dist(H_{1,i}x_1^c, x_i^c) < \epsilon\} . \quad (3)$$

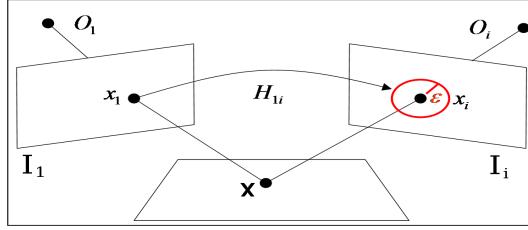


Figure 2: A 3D point X is projected onto points x_1 and x_i on images I_1 and I_i respectively. Point x_1 is said to be repeated by x_i , if a point is detected within a neighborhood of x_i of size ϵ . For planar scenes x_1 and x_i are related by the homography $H_{1,i}$.

Thus, the repeatability rate $r_i(\epsilon)$ of points extracted from image I_i with respect to points from image I_1 is defined by the following equation

$$r_{I_i}(\epsilon) = \frac{|R_{I_i}(\epsilon)|}{\min(\gamma_1, \gamma_i)}, \quad (4)$$

where $\gamma_1 = |\{x_1^c\}|$ and $\gamma_i = |\{x_i^c\}|$ are the total number of points extracted from image I_1 and image I_i respectively. Schmid et al. (2000) conclude that the *Improved-Harris* operator shows the best stability, outperforming every detector included in the survey. Those results have contributed into making the Harris detector the most widely used within CV. However, the list of compared detectors was not exhaustive, examples of those left out include (Beaudet, 1978; Kitchen and Rosenfeld, 1982).

4 Evolving interest point operators with Genetic Programming

The design strategy described in the present work focuses on two main aspects of the detection process: 1) the *structure* of the operator used to compute an interest measure, and 2) the *performance measure* used for evaluation. The structure of every operator presented in Section 3.1 was designed using a particular problem modeling and data representation. Each is derived from a detailed analysis of the observable properties of particular image features that an operator was intended to detect, such as corners, blobs or edges. However, every operator presented thus far can be constructed using a finite set of basic image operations. For instance, every operator employs at least one arithmetic operation, and it also relies on image derivatives and/or Gaussian filtering. From the point of view of measuring the performance of an operator, it has already been discussed that a reliable measure exists, the repeatability rate. Thence, it is reasonable to assume that an *optimal* operator would have the highest repeatability measure on all image sequences where it can be computed. Based on the previous observations the following research question can be stated: *Is it possible to automatically synthesize an interest operator optimized for high performance and constructed from a basic set of image operations?* The hypothesis of this work is that the answer is infact affirmative, and the rest of this section outlines a framework based on GP that accomplishes this task.

The goal of the GP search is to synthesize operators that detect point features that are sparsely distributed, and are invariant to certain transformations. These objectives are not directly related to a particular semantic concept captured within an image. Thence, the GP design process might produce operators that detect features that deviate from the common conceptualization of what an interest point should be, such as a corner. However, this should not be taken as a limitation of the approach. On the

contrary, it is more aptly seen as a possibility to discover the true nature of what an interest point is, or should be, and to better understand what is the most appropriate way to evaluate the detection process.

In order to present the GP-based search two aspects must be defined: 1) the function and terminal sets, which represent the problem's search space; and 2) the evaluation function that defines the fitness landscape.

4.1 Search space

From a careful analysis of the previously described operators it is possible to define a set of primitives that can be used to construct any of them, as well as a large amount of still unknown operators. This leads to the following Function and Terminal sets,

$$\begin{aligned} F = & \{+, |+, |-, |-, |I_{out}|, *, \div, I_{out}^2, \sqrt{I_{out}}, \log_2(I_{out}), EQ(I_{out}), k \cdot I_{out}\} \\ & \cup \left\{ \frac{\delta}{\delta x} G_{\sigma_D}, \frac{\delta}{\delta y} G_{\sigma_D}, G_{\sigma=1}, G_{\sigma=2} \right\}, \\ T = & \{I, L_x, L_{xx}, L_{xy}, L_{yy}, L_y\}, \end{aligned} \quad (5)$$

where I is the input image, and I_{out} can be any of the terminals in T , as well as the output of any of the functions in F ; $EQ(I)$ is an histogram normalization operation; L_u are Gaussian image derivatives along direction u ; G_σ are Gaussian smoothing filters; $\frac{\delta}{\delta u} G_{\sigma_D}$ represents the derivative of a Gaussian function³; and the constant $k = 0.05$.

Note that the authors in (Kitchen and Rosenfeld, 1982; Wang and Brady, 1991; Beaudet, 1978) do not use Gaussian derivatives. However, T is defined in this way because Gaussian derivatives are less susceptible to noise perturbations and provide isotropic properties. It is necessary to remember that every operator constructed with the primitives $p \in \{F \cup T\}$ will only have one true input which is the image I ; and therefore, the differentiation between T and F is only used to help the evolving process by including what is a priori considered useful information (Lin and Bhanu, 2005). It is not claimed that $\{F \cup T\}$ represents an optimal set of primitives for the problem at hand. Nevertheless, from a large set of experimental runs it was confirmed that selective pressure tends to favor the primitives in $\{F \cup T\}$. Other primitives were also tested, such as Gabor filters, Sobel filters, and max and min filters. However, those were consistently removed from the final population and considered unsuitable genetic material; therefore, they are not included in the current discussion.

Figure 3 represents a high level view of the space Ω of possible interest operators constructed with primitives taken from $\{F \cup T\}$. Additionally, a subspace $\Omega_\delta \subset \Omega$ represents the subspace of operators that use image derivatives taken from T to obtain their interest measure. Obviously, image derivatives that were not computed through any of the terminals in T are not considered as part of the genotype. Figure 3 also shows the subspace of operators that rely on measures derived from the local autocorrelation matrix Ω_A . Because of the earlier definition of A , it is affirmed that $\Omega_A \subset \Omega_\delta$. Both Ω_A and Ω_δ contain operators that are similar based on their structure and not their functionality; i.e., based on their genotype and not their phenotype. Another possible subspace $\Omega_\beta \subset \Omega$ contains operators that extract a measure related to surface curvature. In this case, Ω_β contains operators with similar functionality (phenotypical representation). This distinction is made explicitly because a one-to-one correspondence between structure and functionality cannot be expected with this GP approach, what is often called the problem of *competing conventions* (Montana and Davis, 1989). In the current work, this difficulty is made evident when one considers function approximations. For

³All Gaussian filters are applied by convolution.

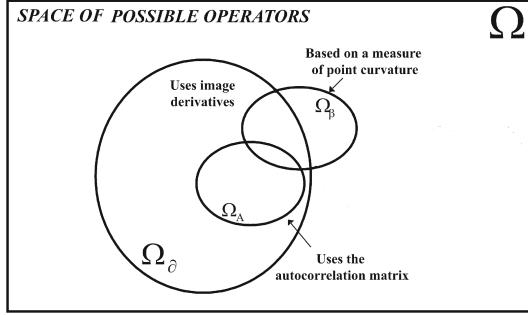


Figure 3: Space of possible interest operators constructed with $\{F \cup T\}$. Three sub-spaces are shown: a) Ω_δ contains operators that use image derivatives taken from T ; b) Ω_A contain operators that employ elements from the autocorrelation matrix A ; and c) Ω_β which contains operators that extract a measure related to surface curvature.

instance, a LoG filter can be approximated by a DoG operation, these operators have very different genotypes in the described GP setting; however, they share a similar phenotypical functionality.

4.2 Fitness evaluation

An appropriate objective function should depend on an operator's repeatability rate $r_{K,J}(\epsilon)$, measured with an operator K and an image sequence J , in such a way that

$$f(K) \propto r_{K,J}(\epsilon). \quad (6)$$

The value $r_{K,J}(\epsilon)$ will represent the performance that K exhibits when it is applied to an image sequence J of progressively transformed images I_i , with $i = 1 \dots N$, starting from the reference image I_1 and extracting a repeatability score $r_{K,I_i}(\epsilon) \forall I_i \in J$ with $i \neq 1$. In this way, $r_{K,J}(\epsilon)$ could be the minimum or maximum value of $\{r_{K,I_i}\}$. In this work, the average repeatability rate obtained on J is employed; a simple but experimentally effective choice (Trujillo and Ologue, 2006a,b),

$$r_{K,J}(\epsilon) = \frac{1}{N-1} \sum_2^N r_{K,I_i}(\epsilon). \quad (7)$$

However, when $r_{K,J}(\epsilon)$ is the only term in $f(K)$ the GP search can get lost in unwanted maxima if appropriate precautions are not taken. For example, in a degenerate case the GP search could exploit individuals that extract useless points clustered together on texture-less regions and still achieve a high repeatability rate; see Figure 4(a). This is undesirable; for example, if the goal of the detection process is to identify unique and distinctive points. Moreover, because the training images employed have texture distributed in a close to uniform manner, it is prudent to predict that a *good* detector will extract uniformly distributed points. Obviously, this constitutes a multi-objective problem, one objective is to promote a high repeatability score and the other is to have a high level of point dispersion or *global separability*. The latter is a property that depends on the specific characteristics of a given image. Nevertheless, it was experimentally determined that including this criterion into the fitness function promotes the emergence of better overall detectors. Therefore, both objectives are combined in a multiplicative

manner,

$$f(K) = r_{K,J}(\epsilon) \cdot \phi_x^\alpha \cdot \phi_y^\beta \cdot N\%^\gamma, \quad (8)$$

where ϕ_u are terms that promote a large *global separability* of detected points. The terms ϕ_u behave like sigmoidal functions, the reason for which is explained below, within a specified interval,

$$\phi_u = \begin{cases} \frac{1}{1 + e^{-a(H_u - c)}}, & \text{when } H_u < H_u^{\max}, \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

Where H_u is the entropy value of the spatial distribution of detected interest points along direction u , on the reference image I_1 of the training set J , given by

$$H_u = - \sum P_j(u) \log_2 [P_j(u)], \quad (10)$$

with $P_j(\cdot)$ approximated by the histogram of interest point localizations using bins of 8×8 pixels. Values for H_x^{\max} and H_y^{\max} are set empirically using the reference image of the training sequence. In order to better explain function ϕ_u an illustrative example is helpful. Figure 4 shows the curve generated by the sigmoidal part of ϕ_x^α with the following parameter values: $\alpha = 20$, $a = 7$ and $c = 5$. In this plot, the position of three individuals on the curve is shown, each has a different amount of *global separability*, or to be more precise, a different entropy value H_x . For instance, the points shown in Figure 4(a) are tightly clustered, thus the associated entropy value $H_x = 5.6$ is low. The next individual is the K_{IPGP2} operator (Trujillo and Olague, 2006a,b), Figure 4(b), which extracts highly repeatable points and maintains a good amount of point dispersion with $H_x = 5.74$. The last individual, Figure 4(c), is an operator that extracts very sparse points and has a low associated $r_{K,J}(\epsilon)$, the individual is therefore biased to the separability part of the fitness function. Experimental results have shown that stable operators with good repeatability tend to have an associated ϕ_x^α located within the bounding box shown in Figure 4. Therefore, values above $H_x^{\max} = 5.9$ are thresholded in order to discourage individuals that only exploit the separability term; using the same analysis, the entropy value along direction y is set to $H_y^{\max} = 5$. The final term,

$$N\% = \left(\frac{\text{extracted points}}{\text{requested points}} \right)^\gamma, \quad (11)$$

is a penalizing factor that reduces the fitness value for detectors that return less than the number of requested points. Because a GP can construct very diverse operators, it is impossible to set a constant value for h in Eq. (1) that would be useful for every operator K . Therefore, a constant number of points are selected, *requested points*=500, for every individual operator in the following manner. After obtaining an interest image I_i^* all the pixels in I_i^* are sorted in descending order and h is set to the value of the 500th highest pixel. Because the condition in Eq. (1) is $K(\mathbf{x}) > h$ and not $K(\mathbf{x}) \geq h$ it is possible that *extracted points* < *requested points*. If the difference is small then the penalization will be negligible. However, for a K that produces an I_i^* with uniform values, such as $K(\mathbf{x}) = 0 \forall \mathbf{x} \in I_i$, then this will result in a high penalty for K because of the inability to produce a discriminate value between interesting and non-interesting pixels. A final heuristic taken in order to promote interest point dispersion was to set the size of the neighborhood \mathbf{W} of Eq. (1) to $n = 5$. This proved to be beneficial during evolution, and n was set to 3 during testing in order to make possible a straightforward comparison with Schmid et al. (2000).

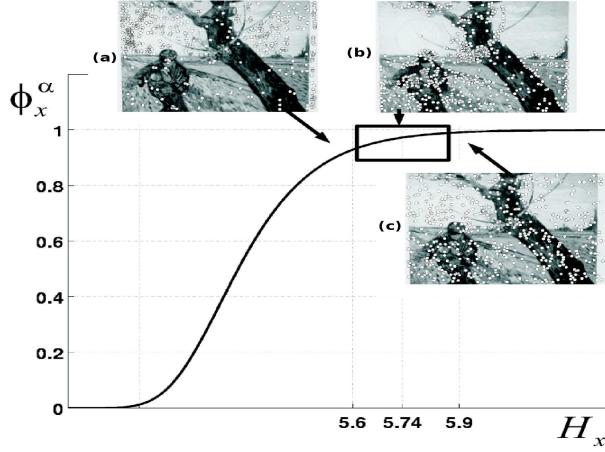


Figure 4: Behavior of the separability term ϕ_x^α of $f(K)$. It shows the associated *separability* value of three operators. From left to right: first, an operator that extracts cluttered points; second, *IPGP2* with a good *separability* of detected points; and finally, an operator biased towards detecting very sparse points.



Figure 5: Samples from the Van Gogh sequence show images: I_1 , I_3 , and I_7 .

5 Experimentation

This section describes the implementation setup, the training and testing sets, previous results, and the new set of operators designs by the proposed GP algorithm.

5.1 Implementation details

The implementation of the previously described approach was programmed on Matlab, with the Genetic Programming toolbox (GPLAB)⁴. The image sequence used for training was the Van Gogh set of a planar scene with rotation transformations, see Figure 5. The Van Gogh sequence has one base image and 16 progressively rotated images, $N = 17$, each with a rotation of 11.25° degrees clockwise from the previous one. However, due to the computational complexity of obtaining interest points from arbitrary operators and calculating a repeatability score for each image, only half of the Van Gogh sequence is used for training. Hence, $N = 9$ with a rotation angle of 22.5° , a trade-off between generality and computational effort. Training and testing sequences were all downloaded from the Visual Geometry Group website⁵, along with Matlab source code that computes the repeatability rate, and binary files for the *Improved Har-*

⁴<http://gplab.sourceforge.net/index.html>, GPLAB A Genetic Programming Toolbox for MATLAB.

⁵Visual Geometry Group: <http://www.robots.ox.ac.uk/~vgg/research/>.

ris detector used for comparison. All experiments were made on a PC with Intel Dual-Core processor and 2Gb of RAM running Linux OS. Table 1 specifies the GP run-time parameters used during the experimental tests. The first three parameters, *population size*, *initialization*, and *genetic operator probabilities*, have canonical values which were set empirically. The next four help to control code bloat. *Tree depth* is dynamically set using two maximum tree depths that limit the size of any given individual within the population. The *dynamic max depth* is a maximum tree depth that may not be surpassed by any individual unless its fitness matches the fitness of the best individual found. When this happens, the *dynamic max depth* is augmented to the tree depth of the new fittest individual. The other, *real max depth* is a hard limit that no individual may surpass under any circumstance. In previous work, (Trujillo and Olague, 2006a,b), tournament selection with lexicographic parsimony pressure was employed. However, this lead to a loss in population diversity without significant code bloat control. Therefore, the new set of results employ *stochastic universal sampling*.

5.2 Preliminary results

Previous work by Trujillo and Olague (2006a,b) presented two operators: K_{IPGP1} and K_{IPGP2} ; where *IPGP* stands for *Interest Point operator with Genetic Programming*.

$$K_{IPGP1}(\mathbf{x}) = G_{\sigma=2} * (G_{\sigma=1} * I - I) , \quad (12)$$

$$K_{IPGP2}(\mathbf{x}) = G_{\sigma=1} * [L_{xx}(\mathbf{x}) \cdot L_{yy}(\mathbf{x}) - L_{xy}^2(\mathbf{x})] . \quad (13)$$

Figure 6 shows an interest image obtained with each operator, along with relevant statistics of their corresponding evolutionary run. The top row are the plots related to K_{IPGP1} found in generation 9; while, the bottom row are plots for K_{IPGP2} found in generation 18. Both experimental runs show low associated diversity, this is mostly related to the tournament selection employed. For K_{IPGP1} the population fitness tends to be higher, a product of the low diversity, and the fact that K_{IPGP1} is situated on a plateau of local maxima. A subspace $\Omega_1 \subset \Omega$ is defined that contains individuals situated within this plateau, these are individuals with similar genetic material to K_{IPGP1} . On the other hand, K_{IPGP2} is obviously located on a more steeper slope of fitness space, evidenced by the more unstable fitness plots of the population. This is further exemplified by the fact that of 30 additional experimental runs none rediscovered the structure of K_{IPGP2} . The mathematical expression of each operator and the corresponding interest image reveal noteworthy properties. K_{IPGP1} can be seen as extracting image borders, applying what can be understood as DoG filtering. K_{IPGP2} , on the other hand, performs a smoothing operation on the determinant of the Hessian,

Parameters	Description and values
<i>Population size</i>	50 individuals.
<i>Generations</i>	50.
<i>Initialization</i>	Ramped Half-and-Half.
<i>Crossover & Mutation prob.</i>	Crossover prob. $p_c = 0.85$; mutation prob. $p_\mu = 0.15$.
<i>Tree depth</i>	Dynamic depth selection.
<i>Dynamic max depth</i>	5 levels.
<i>Real max depth</i>	7 levels.
<i>Selection</i>	Stochastic universal sampling
<i>Survival</i>	Keep best survival strategy.
<i>Fitness function parameters</i>	$a_x = 7, c_x = 5.05, a_y = 6, c_y = 4.3, \alpha = 20, \beta = 20, \gamma = 2$.

Table 1: General parameter settings for our GP framework.

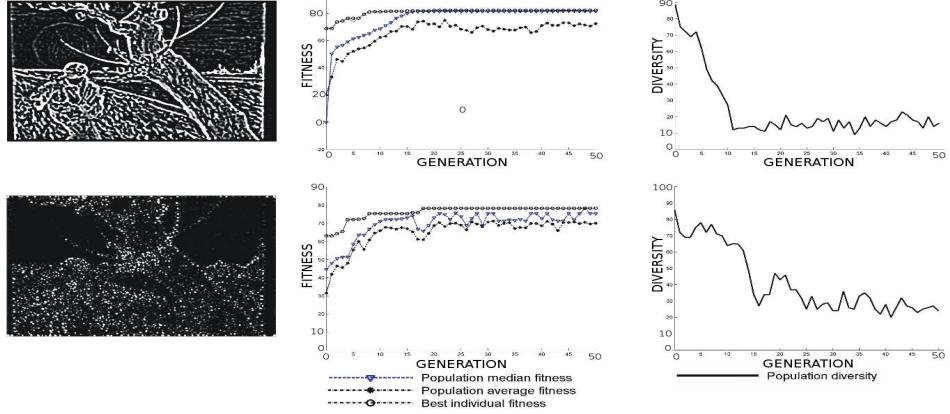


Figure 6: Evolutionary statistics for $IPGP1$ and $IPGP2$, the operators correspond to the first and second row respectively. The columns show from left to right: 1) Sample I^* , 2) Fitness graphs, and 3) Population diversity.

a small modification over the operator originally proposed by Beaudet (1978). Despite the fact that the improvement discovered by K_{IPGP2} is subtle, it is the same type of modification proposed by Schmid et al. (2000) to the Harris operator, what the authors called *Improved Harris*, that produced significant performance gains.

Figure 7 plots the repeatability rate of K_{IPGP1} and K_{IPGP2} on the complete Van Gogh sequence, along with the repeatability rate of the operators proposed in (Beaudet, 1978; Kitchen and Rosenfeld, 1982; Harris and Stephens, 1988; Wang and Brady, 1991; Förstner and Gülich, 1987). For the Harris and Stephens detector the *Improved Harris* is used (Schmid et al., 2000). All the detectors were given the benefit of using isotropic Gaussian derivatives. The plot shows that K_{IPGP1} and K_{IPGP2} outperform all man-made designs, except for the *Improved Harris* operator that has comparable performance. If the additional 3 detectors evaluated by Schmid et al. (2000) are considered, the evolved operators outperform a total of 7 man-made designs. Additionally, the only operators with comparable performance, *Improved Harris*, has a more complex structure which can be a factor when deciding for real-time vision applications.

5.3 Further results

After obtaining the previous results, 30 new runs of the algorithm have yielded the following outcome. Table 2 lists the 15 best solutions found during those executions which are also called the super individuals. From the 30 runs, only 18 generated a solution that can generalize on the complete Van Gogh sequence. Also, from those 18 runs a solution was reproduced on three occasions; thus, the 15 operators presented here. The 12 non-productive runs could have been avoided if the complete Van Gogh sequence was used during training. There is an obvious trade-off between the run-time of the algorithm and the probability of obtaining a useful operator. In order to avoid this bottleneck a larger population size or a grid computing environment could be employed; Lombraña et al. (2007) present preliminary results on this topic. Table 2 shows the name of each operator, the mathematical expression, the number of the corresponding run in which it was obtained, the corresponding subspace, its fitness value, the average repeatability score on the complete Van Gogh sequence, and if the expression was alge-

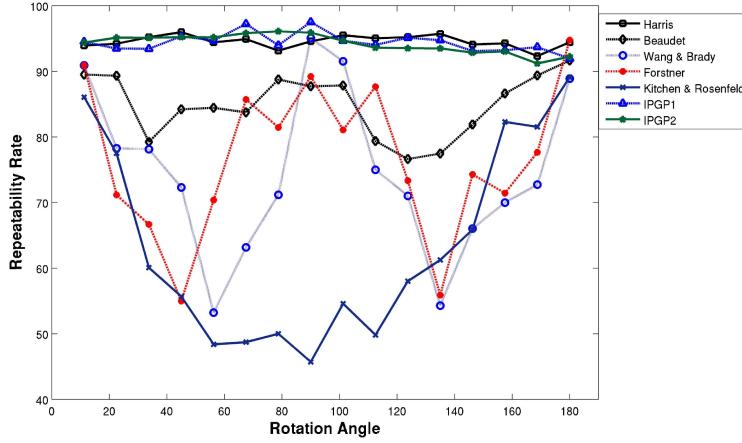


Figure 7: Repeatability score on the complete Van Gogh sequence for K_{IPGP1} and K_{IPGP2} ; plotted for comparison with the operators proposed in (Beaudet, 1978; Kitchen and Rosenfeld, 1982; Harris and Stephens, 1988; Wang and Brady, 1991; Förstner and Gülich, 1987).

Name	Operator	Run	Sub.	Fit.	$r_J *$	Sim.
$IPGP1^*$	$G_2 * I - G_2 * I ^2$	3, 28	Ω_1	77.91	94.78	no
$IPGP3$	$G_1 * G_1 * G_1 * G_2 * G_2 * (G_1 * I - I)$	5	Ω_1	83.98	95.9	no
$IPGP4$	$G_2 * G_2 * G_2 * (G_2 * I - I)$	7	Ω_1	85.98	96.35	no
$IPGP5$	$G_1 * G_2 * I - G_1 * I ^2$	19	Ω_1	83.86	93.22	no
$IPGP6$	$G_2 * G_2 * G_1 * \left(\frac{I}{G_2 * I} \right)$	15	Ω_1	78.13	94.84	no
$IPGP7$	$G_2 * 2 \cdot L_{yy} + 2 \cdot L_{xx} $	4, 9	Ω_∇	78.33	94.92	yes
$IPGP8$	$G_2 * L_{xx} + 2 \cdot G_2(L_{xx} + L_{yy})^2 $	11	Ω_∇	73.86	90.54	yes
$IPGP9$	$G_2 * G_2 * 2 \cdot L_{yy} + 2 \cdot L_{xx} + L_{xy} $	16	Ω_∇	80.3	93.44	no
$IPGP10$	$G_2 * L_{yy} + L_{xx} $	18, 20	Ω_∇	77	92.81	no
$IPGP11$	$G_1 * \left(\frac{G_1 * I}{(G_1 * G_1 * I)^3} \right)$	21	$\Omega_\sigma \setminus \Omega_1$	78.23	92.44	yes
$IPGP12$	$\frac{G_2 * I^{\frac{3}{2}}}{(G_1 * I)^{\frac{9}{4}}}$	22	$\Omega_\sigma \setminus \Omega_1$	72.67	91.91	yes
$IPGP13$	$G_2 * G_2 * [(G_2 * I)(G_2 * G_1 * I - I)]$	6	$\Omega_\sigma \setminus \Omega_1$	85.72	96.37	no
$IPGP14$	$G_2 * G_2 * [(G_2 * G_2 * G_2 * I)(G_2 * G_2 * I - I)]$	23	$\Omega_\sigma \setminus \Omega_1$	85.94	96.5	no
$IPGP15$	$G_2 * [G_2 * G_2 * I - G_1 * I - G_2 * G_2 * I]$	24	$\Omega_\sigma \setminus \Omega_1$	85.81	95.15	no
$IPGP16$	$G_2 * G_2 * I$ $G_2 * G_2 * [G_1 * G_2 * I^2 - I^2]$ $G_2 * G_2 * G_1 * I$	30	$\Omega_\sigma \setminus \Omega_1$	84.63	96.8	no

* To obtain r_J the size of the neighborhood \mathbf{W} for non-maximum suppression was set to $n = 3$.

Table 2: Evolved interest operators.

braically simplified. Table 2 is organized into three sets, each presenting operators from the subspaces Ω_1 , $\Omega_\sigma \setminus \Omega_1$ and Ω_∇ . The subspace $\Omega_\sigma \subset \Omega$ contains operators that do not include Gaussian derivatives L_u as part of their genetic material; i.e., they rely on Gaussian filters and arithmetic operations. Next, $\Omega_1 \subset \Omega_\sigma$, as mentioned above, contains operators that are genetically similar to K_{IPGP1} . Some of the GP runs converged around the local optima in which $IPGP1$ is located. Finally, $\Omega_\nabla \subset \Omega_\delta$ contains operators that explicitly apply the Laplacian operation to the image, making it an obvious subspace of Ω_δ . Notice that Ω_∇ and Ω_1 are located in different regions of Ω defined by the genotype, and at the same time share the same phenotypical representation caused by the competing conventions problem. Additionally, the GP also finds unorthodox

operators that are not commonly used in vision applications; each is described next.

5.4 Discussion and comparisons

The discussion begins by focusing on operators within Ω_1 . First, special attention is given to K_{IPGP1^*} because of its extremely simple structure, a particularity it shares with K_{IPGP1} , and hence the name assigned to it. K_{IPGP3} and K_{IPGP4} , on the other hand, are proportional to K_{IPGP1} ; they apply the same filtering of the form $G_{\sigma_1} * I - G_{\sigma_2} * I$ with $\sigma_1 > \sigma_2$. The same analogy can be made between K_{IPGP1^*} and K_{IPGP5} , because the only difference between them is the amount of image smoothing used. The additional smoothing operations are related with the theory of scale-space analysis, where larger image structures are salient at larger smoothing scales (Lindeberg, 1998). It is evident that the GP process focuses the detection strategy on larger image structures which are indeed more stable than smaller features. This particular result exhibits the resourcefulness of EC methods.

Operators that are proportional to K_{IPGP1} will detect corners, edges and blobs of salient low intensity regions; the additive inverse of K_{IPGP1} extracts salient high intensity regions. It is interesting to note that the basic structure of K_{IPGP1} has been rediscovered a total of three times by the GP search, while its additive inverse has yet to appear (Trujillo and Olague, 2007). This is probably related to the training sequence used and not due to any bias within the search process. K_{IPGP1^*} on the other hand, because of its absolute value, identifies maxima related to both K_{IPGP1} and its additive inverse K_{-IPGP1} . An interesting property of the operators from Ω_1 is that all employ a DoG filter.

Proposition 1: Both K_{IPGP1} and K_{IPGP1^*} are proportional to DoG (Difference-of-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

$$G_{\sigma} * I - I = G_{\sigma} * G_{\hat{\sigma}} * \hat{I} - G_{\hat{\sigma}} * \hat{I} \propto G_{\sigma + \hat{\sigma}} * \hat{I} - G_{\hat{\sigma}} * \hat{I} = (G_{\sigma + \hat{\sigma}} - G_{\hat{\sigma}}) * \hat{I}. \quad (14)$$

Given Proposition 1, K_{IPGP1^*} , K_{IPGP3} , K_{IPGP4} , and K_{IPGP5} , are all related by their use of a DoG filter, not just by their genetic similarities with K_{IPGP1} . The final operator assigned to Ω_1 is K_{IPGP6} , its main difference to K_{IPGP1} is the fact that in place of an image subtraction it uses a protected image division. This makes K_{IPGP6} a very close neighbor of K_{IPGP1} in terms of its genetic content, but also quite different considering the operation it performs.

The next group of operators in Table 2 are those from Ω_∇ . Operators K_{IPGP7} and K_{IPGP10} perform a similar Laplacian operation, differing only by a dismissible constant factor, which means that the GP converged to this precise solution four times. It is easy for the GP search, thanks to the terminal set, to find an above average fitness region using the provided Gaussian derivatives. Furthermore, K_{IPGP8} and K_{IPGP9} perform more complex operations, but still rely on the Laplacian to obtain their interest measure. For example, K_{IPGP9} is the sum of all second order derivatives, while K_{IPGP8} uses the square of the Laplacian. Hence, because the DoG filter is an approximation of the Laplacian, Ω_∇ and Ω_1 contain operators with competing conventions.

Finally, Table 2 presents six operators from $\Omega_\sigma \setminus \Omega_1$; they apply interesting variations of the DoG filtering performed by operators in Ω_1 , such as a DoG filter applied to the square of an image in K_{IPGP12} , K_{IPGP14} and K_{IPGP16} . The protected division was used by three of the individuals, only 4 out of the 18 super individuals synthesized use this primitive. Ratios of two measures appear to be more useful when Gaussian derivation is not employed. Figures 8, 9 and 10 present the evolution statistics of each

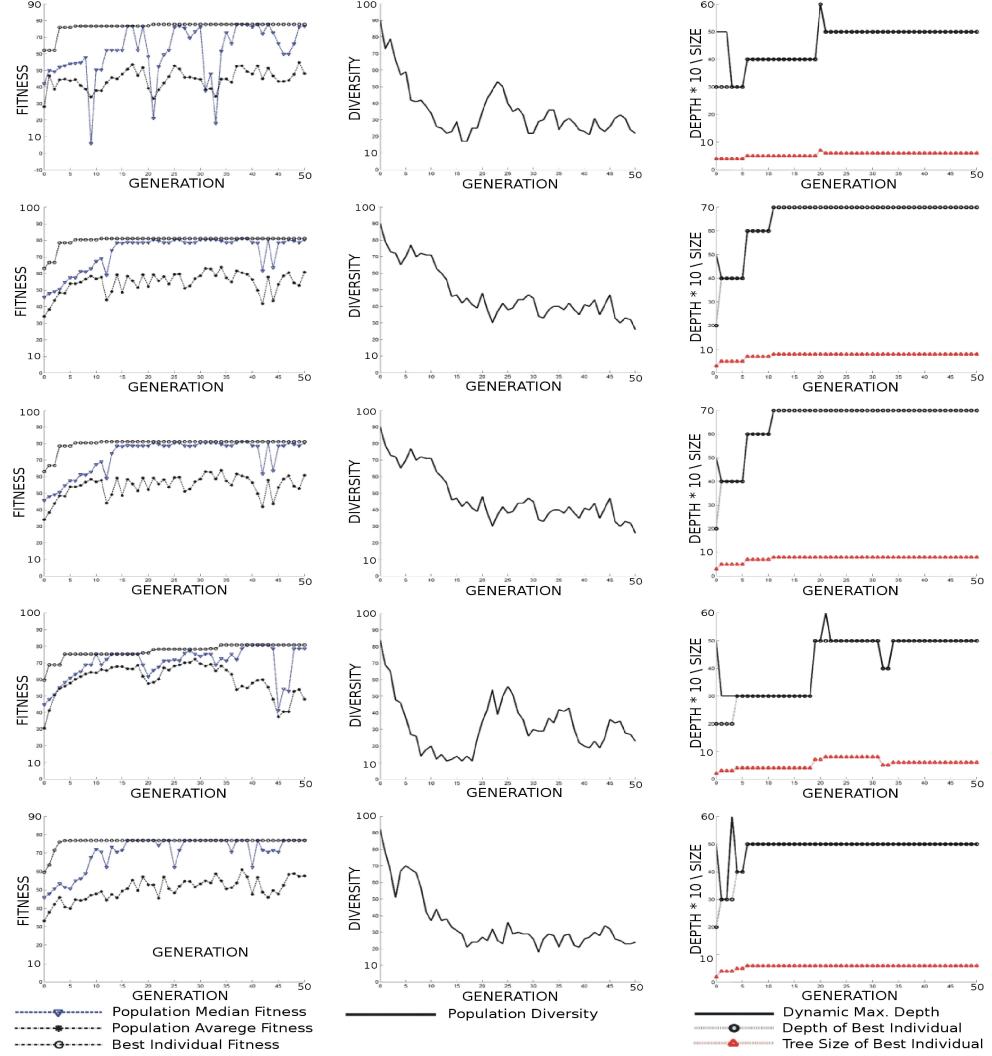


Figure 8: Evolution statistics for operators presented in Table 2 are presented by columns from left to right: 1) fitness (median, average, and best fitness), 2) population diversity, 3) complexity (*dynamic max depth*, best individual tree depth and number of nodes); by rows, from top to bottom: *IPGP1**, *IPGP3*, *IPGP4*, *IPGP5*, and *IPGP6 (continued)*.

operator described in Table 2, presented in the same order row by row. Columns, from left to right, are for fitness statistics, population diversity, and complexity of best individual. Fitness plots include the median population fitness, the average population fitness and the fitness of the best individual found thus far in every generation. Diversity graphs represent the amount of genetic diversity found in the population at each generation. Finally, the complexity graphs describe the behavior of the *dynamic maximum tree depth* parameter, along with the tree depth of the fittest individual (both of these values multiplied by 10 to facilitate viewing), and the number of nodes of the fittest

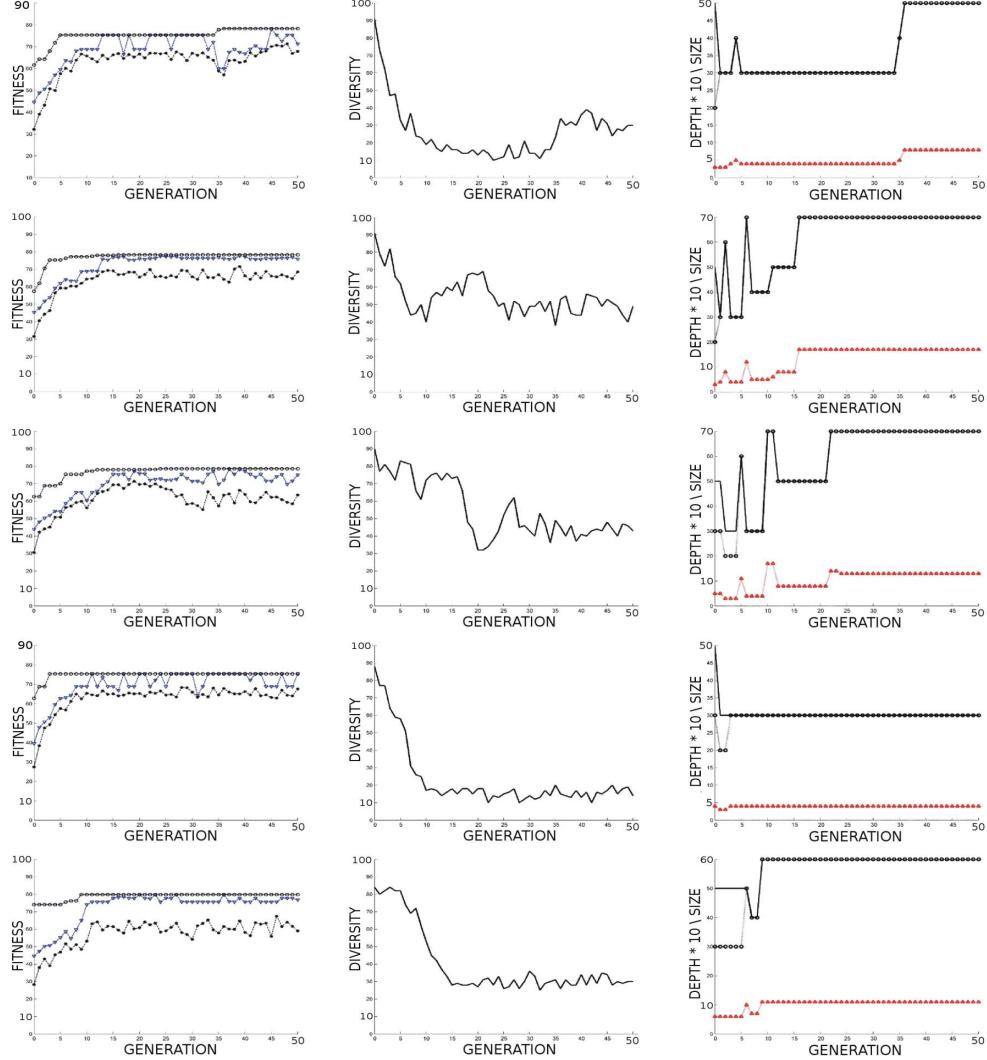


Figure 9: Evolution statistics for operators presented in Table 2, by rows: *IPGP7*, *IPGP8*, *IPGP9*, *IPGP10*, and *IPGP11*.

individual. For operators discovered in more than one run, the statistics of a single run are used to present the behavior of the evolutionary process. Most of the operators are found early in the GP search, before generation 25, with K_{IPGP5} , K_{IPGP7} , K_{IPGP12} K_{IPGP14} , K_{IPGP15} , and K_{IPGP16} being the exceptions. Operators from $\Omega_\sigma \setminus \Omega_1$ are mostly found in later generations, probably because their unorthodox measures require more exploration. Furthermore, statistics show that populations that converge to $\Omega_\sigma \setminus \Omega_1$ have lower median and average fitness, with low median peaks even near the end of the run. Conversely, populations that converged to Ω_∇ have higher performing populations, where image derivatives are able to capture, in a direct manner, the high variation property that every interest point requires. The diversity plots of all operators are closely related to the complexity of the best individual found. The smaller the super

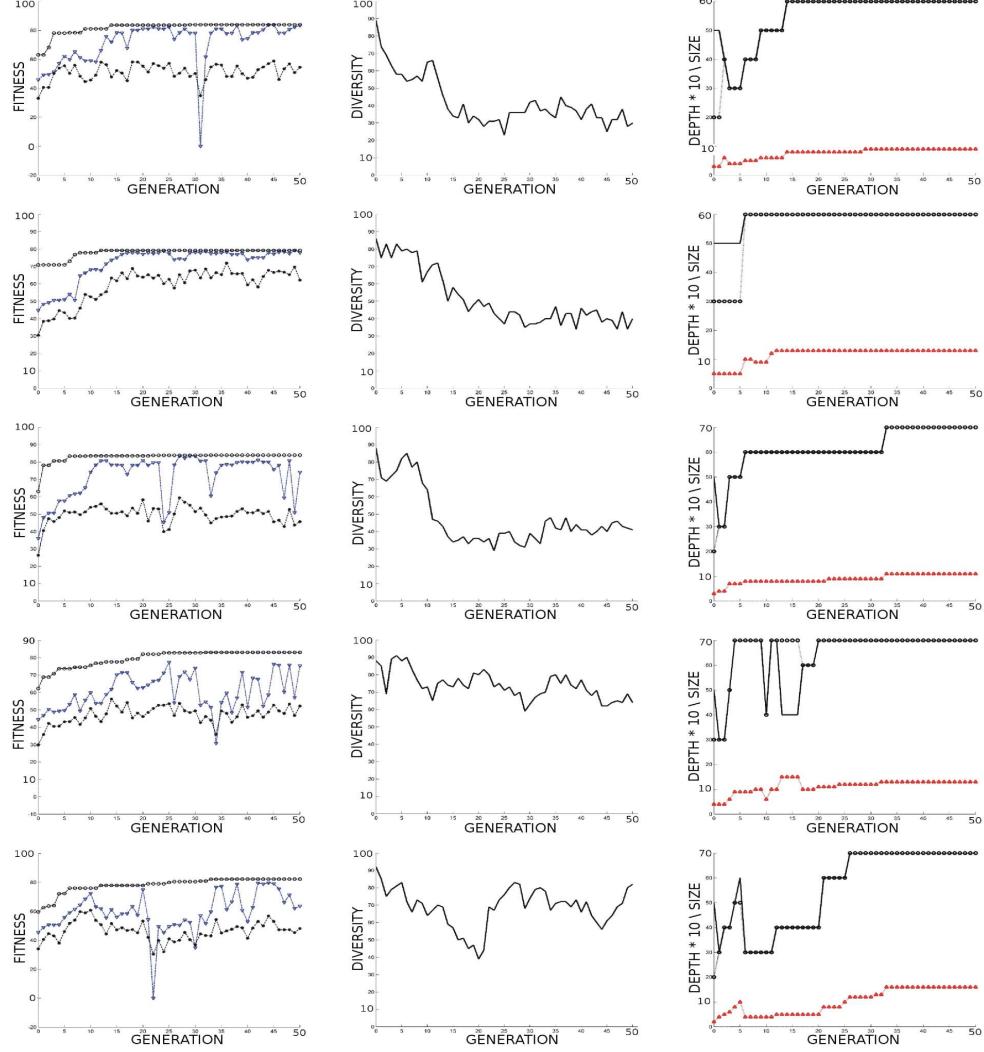


Figure 10: Evolution statistics for operators presented in Table 2, by rows: *IPGP12*, *IPGP13*, *IPGP14*, *IPGP15*, and *IPGP16*.

individual, in terms of number of nodes, the less diverse the population; the opposite being true for larger super individuals. Therefore, populations that converge towards Ω_1 produce the best solution in earlier generations, while populations that converge towards $\Omega_\sigma \setminus \Omega_1$ require more evolution time. The complexity plots also show the benefits of the *dynamic max depth* parameter. Early generations try out simple operators first, and the complexity is allowed to gradually increase only when an associated fitness improvement is found. One final observation regarding the operators presented in Table 2 is related to the filtering of some genetic primitives by selective pressure. From the 20 primitives in $\{F \cup T\}$, six of them never appeared in any of the super individuals, including \sqrt{I} , \log_2 , $\frac{I}{2}$, $EQ(I)$, and more notably, the first order Gaussian derivatives L_u despite their wide ranging use in corner and edge detection.

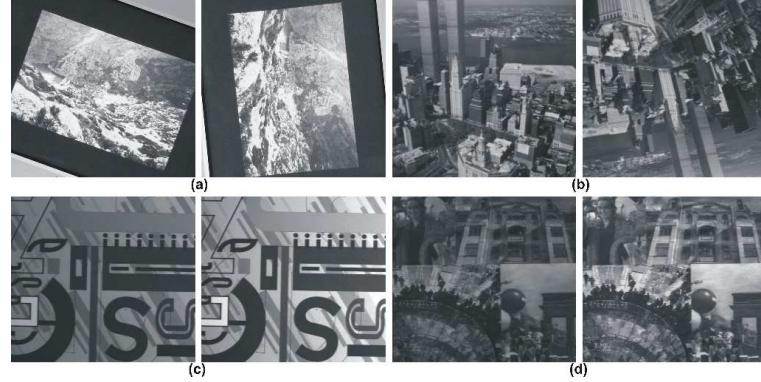


Figure 11: The reference image (right) and a transformed image (left) for each testing sequence. a) Mars ($N = 18$, rotation), b) New York ($N = 35$, rotation), c) Graph ($N = 12$, illumination), d) Mosaic ($N = 10$, illumination).

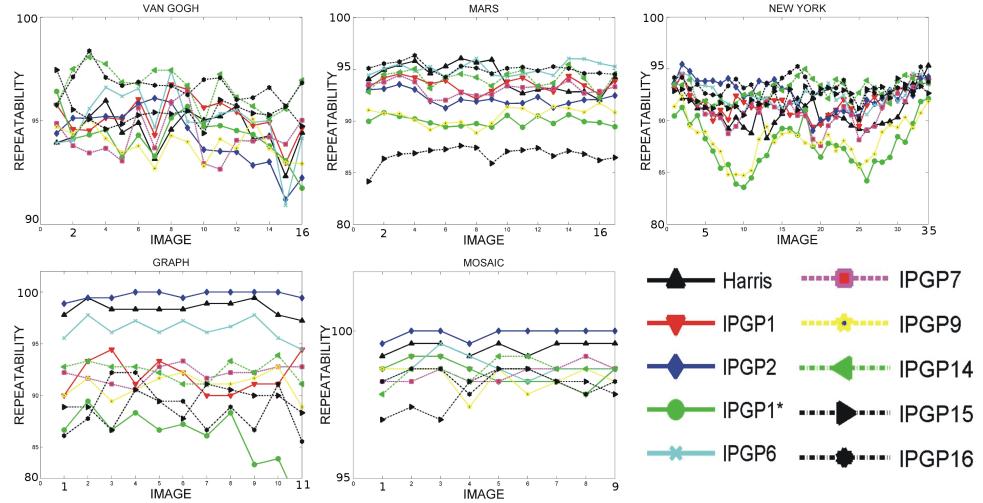


Figure 12: Repeatability for each operator, computed for each image sequence.

From Table 2, a set of 7 operators were selected for further testing along with K_{IPGP1} , K_{IPGP2} , and the *Improved Harris* detector. The operators chosen are: K_{IPGP1^*} , for its simplicity and complementary structure to K_{IPGP1} ; K_{IPGP6} for its unique operation; K_{IPGP7} that represents the 4 operators that use the Laplacian explicitly in Ω_∇ ; K_{IPGP9} , the sum of second order derivatives; and K_{IPGP14} , K_{IPGP15} , and K_{IPGP16} for their unorthodox measures. Other operators are not presented in further tests because some are redundant, and others achieve lower fitness scores. Figure 11 presents 4 image sequences used for testing, where Mars and New York have rotation transformations, and Graph and Mosaic have illumination changes. Considering the fact that the sequence used for training only presents rotation transformations, there is no guarantee that the evolved functions will show stability under other transformations. Furthermore, the Van Gogh sequence is a textured scene while New York has a more

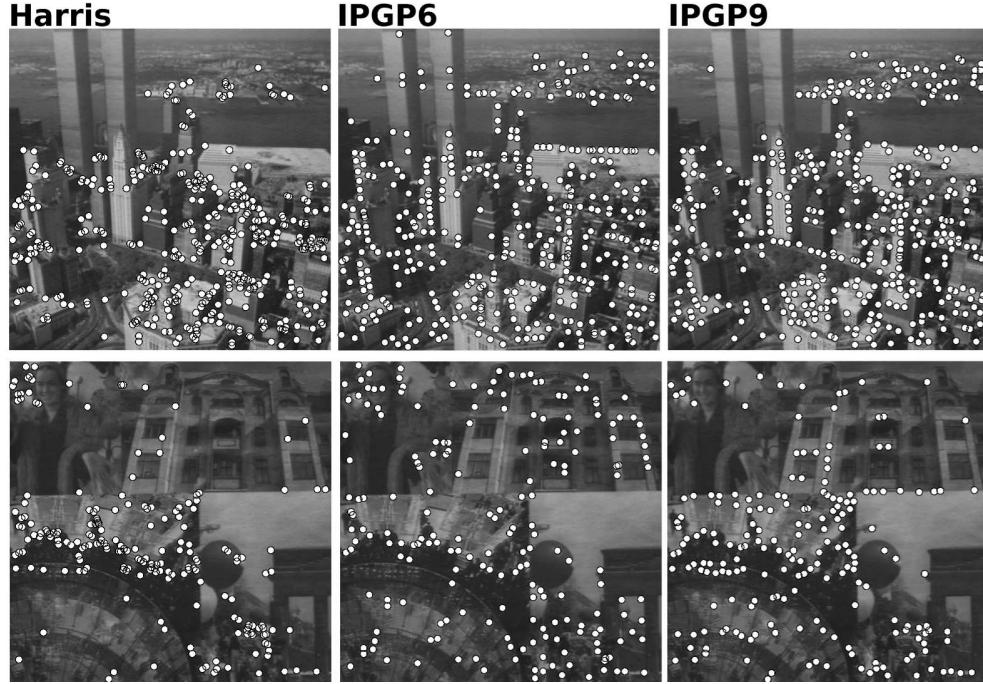


Figure 13: A qualitative comparison showing points with the highest interest measure. First row, New York image with 400 points. Second row, Mosaic image with 200 points.

geometric structure, this could be another possible source of problems.

Figure 12 shows the performance plots. First, for the Van Gogh sequence, as expected, all operators have a high and stable performance, note the scale. For the Mars sequence, high performance is also obtained by most of them, which is an expected result considering the similarities with the training sequence. Only K_{IPGP15} is decidedly worse than the rest, while K_{IPGP1^*} and K_{IPGP9} are only slightly below the average. Encouraging results are obtained for the New York sequences, given the differences between it and the training set. Nevertheless, all operators were able to generalize to the different type of scene, with K_{IPGP1^*} and K_{IPGP9} only slightly below average.

For the sequences that present illumination transformations, the detection method was necessarily modified. Specifically, a fixed number of points cannot be extracted from every image in the sequence because the intensity values are altered. Therefore, the number of extracted points is equal to the number of points detected by the *Improved Harris* detector. Future work requires an optimal threshold to be determined for each operator; however, the solution used is sufficient to illustrate the performance of the described operators. Surprisingly, on both illumination sequences every detector obtained a high performance score, with K_{IPGP2} consistently receiving the slightly higher measures. This shows that the GP search has located operators that can generalize to different types of tests, and match the state-of-the-art in interest point detection. It could be argued that the evolved operators are only capable of mimicking the performance of the Harris detector due to the performance similarity. However, a qualitative comparison shows that the operators found by the GP are capable of achieving a high fitness while also focusing on different image features, see Figure 13.

6 Summary and concluding remarks

This paper presented a GP-based methodology that automatically synthesizes operators that detect interest points on images. The problem formulation poses the design process of an interest point operator as an optimization/search problem, and employs simulated evolution to solve it. A careful analysis of previously proposed detectors revealed that a finite set of simple image operations can be used to define a GP's function and terminal sets. However, the manner in which those elements are combined is not a trivial task. Moreover, close attention was given to define an appropriate objective function. A well-established performance measure was chosen as the main component of the fitness measure, the repeatability rate. In addition, the objective function also promotes a good dispersion of detected interest points. Therefore, operators synthesized with this approach reveal reasonable and desirable properties that make them suitable for many types of vision applications that rely on the detection of salient or interesting image regions. The experimental results presented a total of 17 operators, including the two obtained in earlier work (Trujillo and Olague, 2006a,b).

Strong experimental evidence suggest that the GP-based approach synthesizes robust and effective interest point detectors. A noteworthy result is the fact that the operators are able to generalize quite well to different types of scenes and image transformations. Furthermore, the operators outperform all previously man-made designs except for the *Improved Harris* detector which achieved similar performance. However, the GP search did produce operators that are functionally unique and capable of identifying qualitatively different interest points, when applied to the same image, with a high performance.

On the other hand, interesting properties could be observed when the convergence tendencies of the GP search were analysed. For instance, the fact that the algorithm mostly produced operators that performed LoG or DoG filtering in their computation. The GP basically promotes the detection of edge or blob points and not necessarily image corners. However, this was not always the case because the GP also produced some less obvious and unorthodox operators that were still highly competitive.

Future work on this research will focus on the evaluation scheme. Here, the problem was formulated as a mono-objective optimization process where two performance criteria were combined: the repeatability rate and the amount of global separability. However, there is no prior reason to assert that this is the correct approach, and a truly multi-objective approach could be more insightful and productive. Thence, through multi-objective evolutionary algorithms the problem can be restated and additional objectives could be added in a principled manner.

Acknowledgements

Research funded by UC MEXUS-CONACyT, the Ministerio de Educación y Ciencia (project Oplink - TIN2005-08818-C04), the LAFMI project, and the Junta de Extremadura Spain. First author supported by scholarship 174785 from CONACyT México. Dr. Olague would like to thank Extremadura regional government for the grant received while he was on his sabbatical leave at the University of Extremadura.

References

- Asada, H. and Brady, M. (1986). The curvature primal sketch. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(1):2–14.
- Beaudet, P. R. (1978). Rotational invariant image operators. In *Proceedings of the 4th International Joint Conference on Pattern Recognition (ICPR)*, Tokyo, Japan, pages 579–583.
- Cagnoni, S., Lutton, E., and Olague, G. (Eds.) (2008). *Genetic and Evolutionary Computation for Evolutionary Computation* Volume x, Number x

- Image Processing and Analysis*, volume 8 of *EURASIP Book Series on Signal Processing and Communications*. Hindawi Publishing Corporation.
- Ebner, M. (1998). On the evolution of interest operators using genetic programming. In Poli, R. et al., editors, *Late Breaking Papers at EuroGP'98*, pages 6–10. The University of Birmingham, UK.
- Ebner, M. and Zell, A. (1999). Evolving task specific image operator. In Poli, R. et al., editors, *First European Workshops, EvolISP'99 and EuroEcTel'99, Göteborg, Sweden, May 26-27*, volume 1596 of *LNCS*, pages 74–89.
- Förstner, W. and Gülich, E. (1987). A fast operator for detection and precise location of distinct points, corners and centres of circular features. In *ISPRS Intercommission Conference on fast processing of photogrammetric data*, pages 149–155.
- Harris, C. and Stephens, M. (1988). A combined corner and edge detector. In *Proceedings from the Fourth Alvey Vision Conference*, volume 15, pages 147–151.
- Hernández, B., Olague, G., Hammoud, R., Trujillo, L., and Romero, E. (2007). Visual learning of texture descriptors for facial expression recognition in thermal imagery. *Computer Vision and Image Understanding, Special Issue on Vision Beyond the Visual Spectrum*, 106(2-3):258–269.
- Ke, Y. and Sukthankar, R. (2004). Pca-sift: A more distinctive representation for local image descriptors. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 27 June - 2 July, Washington, DC*, volume 2, pages 506–513. IEEE Computer Society.
- Kitchen, L. and Rosenfeld, A. (1982). Gray-level corner detection. *Pattern Recognition Letters*, 1:95–102.
- Laptev, I. and Lindeberg, T. (2003). Space-time interest points. In *Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV)*, page 432, Washington, DC. IEEE Computer Society.
- Lin, Y. and Bhanu, B. (2005). Evolutionary feature synthesis for object recognition. *IEEE Transactions on Systems, Man and Cybernetics, Part C, Special Issue on Knowledge Extraction and Incorporation in Evolutionary Computation*, 35(2):156–171.
- Lindeberg, T. (1998). Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):79–116.
- Lombraña, D., de Vega, F. F., Olague, G., Trujillo, L., and Segal, B. (2007). Customizable execution environments with virtual desktop grid computing. In *PDCS 2007: Proceedings of the 19th IASTED International Conference on Parallel and Distributed Computing Systems, 19-21 November, Cambridge MA, USA*. Morgan Kaufmann Publishers Inc.
- Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Proceedings of the International Conference on Computer Vision (ICCV), 20-25 September, Kerkyra, Corfu, Greece*, volume 2, pages 1150–1157. IEEE Computer Society.
- Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., and Gool, L. V. (2005). A comparison of affine region detectors. *International Journal of Computer Vision*, 65(1-2):43–72.
- Montana, D. J. and Davis, L. (1989). Training feedforward neural networks using genetic algorithms. In Sridharan, S., editor, *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, pages 762–767, San Francisco, CA. Morgan Kaufman.
- Moravec, H. P. (1977). Towards automatic visual obstacle avoidance. In *IJCAI*, page 584.
- Olague, G., Cagnoni, S., and Lutton, E. (2006a). Preface: introduction to the special issue on evolutionary computer vision and image understanding. *Pattern Recognition Letters*, 27(11):1161–1163.

- Olague, G., Fernández, F., Pérez, C. B., and Lutton, E. (2006b). The infection algorithm: An artificial epidemic approach for dense stereo correspondence. *Artificial Life*, 12(4):593–615.
- Olague, G. and Hernández, B. (2005). A new accurate and flexible model based multi-corner detector for measurement and recognition. *Pattern Recognition Letters*, 26(1):27–41.
- Olague, G. and Mohr, R. (2002). Optimal camera placement for accurate reconstruction. *Pattern Recognition*, 35(4):927–944.
- Olague, G. and Puente, C. (2006). The honeybee search algorithm for three-dimensional reconstruction. In Rothlauf, F. et al., editors, *Proceedings from EvoWorkshops 2006 (EVOIASP), Budapest, Hungary, April 10-12*, volume 3907 of *LNCS*, pages 427–437. Springer-Verlag.
- Poli, R. (1996). Genetic programming for feature detection and image segmentation. In Fogarty, T., editor, *Proceedings of the AISB'96 Workshop on Evolutionary Computation, April, Brighton, UK*, volume 1143 of *LNCS*, pages 110–125. Springer-Verlag.
- Poli, R., Langdon, W. B., and McPhee, N. F. (2008). *A field guide to genetic programming*. Published via <http://lulu.com> and freely available at <http://www.gp-field-guide.org.uk>. (With contributions by J. R. Koza).
- Schmid, C. and Mohr, R. (1997). Local grayvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5):530–534.
- Schmid, C., Mohr, R., and Bauckhage, C. (2000). Evaluation of interest point detectors. *International Journal of Computer Vision*, 37(2):151–172.
- Shi, J. and Tomasi, C. (1994). Good features to track. In *Proceedings of the 1994 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June, Seattle, WA, pages 593–600. IEEE Computer Society.
- Sun, Z., Bebis, G., and Miller, R. (2004). Object detection using feature subset selection. *Pattern Recognition*, 37(11):2165–2176.
- Trujillo, L. and Olague, G. (2006a). Synthesis of interest point detectors through genetic programming. In Cattolico, M., editor, *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO), Seattle, WA, July 8-12*, volume 1, pages 887–894. ACM.
- Trujillo, L. and Olague, G. (2006b). Using evolution to learn how to perform interest point detection. In *Proceedings of the 18th International Conference on Pattern Recognition (ICPR)*, 20-24 August, Hong Kong, China, volume 1, pages 211–214. IEEE Computer Society.
- Trujillo, L. and Olague, G. (2007). Scale invariance for evolved interest operators. In *Proceedings of EvoIASP2007 Ninth European Workshop on Evolutionary Computation in Image Analysis and Signal Processing, incorporated in Evo* 2007, 11-13 April, Valencia, Spain*, LNCS, pages 423–430. Springer-Verlag.
- Trujillo, L., Olague, G., de Vega, F. F., and Lutton, E. (2007a). Evolutionary feature selection for probabilistic object recognition, novel object detection and object saliency estimation using gmms. In *Proceedings from the 18th British Machine Vision Conference, 10-13 September, Warwick, UK*. British Machine Vision Association.
- Trujillo, L., Olague, G., Legrand, P., and Lutton, E. (2007b). Regularity based descriptor computed from local image oscillations. *Optics Express*, 15:6140–6145.
- van de Weijer, J., Gevers, T., and Bagdanov, A. D. (2006). Boosting color saliency in image feature detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(1):150–156.
- Wang, H. and Brady, J. (1991). Corner detection for 3d vision using array processors. In *Proceedings from BARNAIMAGE 91, Barcelona, Spain*, Secaucus, NJ. Springer-Verlag.
- Zhang, M., Ciesielski, V. B., and Andreae, P. (2003). A domain-independent window approach to multiclass object detection using genetic programming. *EURASIP Journal on Applied Signal Processing, Special Issue on Genetic and Evolutionary Computation for Signal Processing and Image Analysis*, 2003(8):841–859.