# Multiobjective Design of Operators that Detect Points of Interest in Images

Leonardo Trujillo,
Gustavo Olague
EvoVisión Project,
CICESE Research Center,
Ensenada, B.C., México.
[trujillo,olague]@cicese.mx

Evelyne Lutton
APIS Team, INRIA-Futurs,
Parc Orsay Université 4,
ORSAY Cedex, France.
evelyne.lutton@inria.fr

Francisco Fernández de
Vega
Grupo de Evolución Artificial,
Universidad de Extremadura,
Mérida, Spain.
fcofdez@unex.es

## ABSTRACT

In this paper, a multiobjective (MO) learning approach to image feature extraction is described, where Pareto-optimal interest point (IP) detectors are synthesized using genetic programming (GP). IPs are image pixels that are unique, robust to changes during image acquisition, and convey highly descriptive information. Detecting such features is ubiquitous to many vision applications, e.g. object recognition, image indexing, stereo vision, and content based image retrieval. In this work, candidate IP operators are automatically synthesized by the GP process using simple image operations and arithmetic functions. Three experimental optimization criteria are considered: 1) the repeatability rate; 2) the amount of global separability between IPs; and 3) the information content captured by the set of detected IPs. The MO-GP search considers Pareto dominance relations between candidate operators, a perspective that has not been contemplated in previous research devoted to this problem. The experimental results suggest that IP detection is an ill-posed problem for which a single globally optimum solution does not exist. We conclude that the evolved operators outperform and dominate, in the Pareto sense, all previously man-made designs.

## Categories and Subject Descriptors

I.4.7 [**Image Processing and Computer Vision**]: Feature Measurement—*feature representation, invariants*; I.2.2 [**Artificial Intelligence**]: Automatic Programming—*program synthesis*

## General Terms

Algorithms, Experimentation, Performance, Theory.

## Keywords

Multiobjective optimization, Interest point detection

## 1. INTRODUCTION AND MOTIVATION

Currently, many computer vision systems rely on the detection of stable and informative features on digital images [7, 15, 22, 24].

The simplest features are known as *interest points* (IP), also referred to as *singular points* or *key points*. Researchers normally design specialized operators to detect IPs [11, 12, 20], or choose one of the detectors currently available within computer vision literature [4, 29]. Nevertheless, designing an operator is not a trivial task, and selecting between previously proposed methods can be overwhelming due to the large number of proposed methods. This paper presents a Genetic Programming (GP) approach that automatically synthesizes several IP detectors using a multiobjective (MO) problem formulation. Here, eight operators are presented, each designed using the MO-GP approach based on Pareto optimality. As a result, it is possible to articulate novel insights regarding this well-known computer vision problem; such as the apparent conflict between reasonable evaluation criteria, and the inferior performance of man-made designs when compared with the experimental Pareto-optimal set.

The remainder of this paper proceeds as follows: Section 2 gives an introduction to the IP detection problem, reviews previously proposed operators, discusses how IP detection can be evaluated and introduces the proposal of the current work. Afterwards, three performance criteria that can be experimentally evaluated are defined in Section 3. Section 4 presents a brief overview of MO optimization. Section 5 discusses how IP detectors can be synthesized through GP, and details the MO proposal. Finally, experimental results are presented in Section 6 and concluding remarks are given in Section 7.

## 2. INTEREST POINT DETECTION

Historically, work devoted to IP detection is a result of computer vision research devoted to the corner detection problem. A common taxonomy of corner detection methods employs three classes: *contour based methods* [1], *parametric model based methods* [21, 17] and *intensity based methods* [16, 2, 8, 6, 28, 5, 23]. The class of corner detectors that operate directly on the intensity image are more appropriately referred to as *interest point detectors*, a subtle difference. Corners are point features located at line and surface junctions (e.g. L, T, Y and X junctions) [17, 24]. On the other hand, IPs may include such features, as well as others that are less obvious to human interpretation. A measure of how *interesting* an image region is, can be computed using a specially designed operator $K : \Re^+ \rightarrow \Re$, where different region detectors employ different operators. A *detector* refers to the complete algorithmic process that extracts interest points. On the other hand, an *operator* only computes the interest measure for each image pixel. Applying $K$ to an image $I$ yields what can be understood as an *interest image* $I^*$. Afterwards, most detectors follow the same basic process: $a)$ non-maxima suppression that eliminates pixels that are not lo-

cal maxima; and *b*) a thresholding step that obtains the final set of points; see Figure 1. Both of the final steps are tuned empirically, a procedure also applied in the current work.

*Definition 1.* Let $K$ be an image operator that is applied to an image $I$, thus generating an interest image $K(I) = I^*$. Then, a pixel $\mathbf{x} \in I$ is tagged as an IP if the following conditions hold,

$$[K(\mathbf{x}) > \max\{K(\mathbf{x_W})|\forall \mathbf{x_W} \in \mathbf{W}, \mathbf{x_W} \neq \mathbf{x}\}] \wedge [K(\mathbf{x}) > h],$$
(1)

where $\mathbf{W}$ is a neighborhood of size $n \times n$ around $\mathbf{x}$ and $h \in \mathbb{R}$ is a threshold.

The first condition in Eq. (1) accounts for non-maximum suppression and the second is the thresholding step; experiments in this work use $n = 5$ while $h$ is operator dependent. The threshold $h$ cannot be set a priori because the GP can design very different kinds of operators. Therefore, a maximum of 500 IPs are selected from an image for every operator $K$ produced by the GP process. 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.

Several improvements to the described process have been proposed, including a space-time analysis [10] and detection that accounts for color information [27]. Color and temporal information are not addressed in the present work; however, doing so is not beyond the scope of the proposed approach.

## 2.1 Popular IP operators

Some detectors base their interest measure $K$ on the local autocorrelation matrix $A$, which characterizes the gradient distribution around each image pixel,

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

where $\sigma_D$ and $\sigma_I$ 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

$$L_u(\mathbf{x}, \sigma_D) = \frac{\delta}{\delta u} G_{\sigma_D} * I(\mathbf{x}),$$
(3)

$G_\sigma$ is a Gaussian smoothing function with standard deviation $\sigma$; $\sigma_D = 1$ is used unless noted otherwise. Detectors based on $A$ include those proposed by Harris and Stephens [6], Förstner [5], and Shi and Tomasi [23] with their corresponding interest measures given by

$$K_{Harris\&Stephens}(\mathbf{x}) = det(A) - k \cdot Tr(A)^2,$$
(4)

$$K_{Forstner}(\mathbf{x}) = \frac{det(A)}{Tr(A)},$$
(5)

$$K_{Shi\&Tomasi}(\mathbf{x}) = \min\{\lambda_1, \lambda_2\},$$
(6)

where $\lambda_1, \lambda_2$ are the two eigenvalues of $A$. The definition of $A$ is taken from [22], used with the *Improved Harris* detector.

Beaudet [2] proposed the determinant of the Hessian, that is proportional to the Gaussian surface curvature, as an interest measure

$$K_{Beaudet}(\mathbf{x}) = I_{xx}(\mathbf{x}) \cdot I_{yy}(\mathbf{x}) - I_{xy}^2(\mathbf{x}),$$
(7)

where $I_u(\mathbf{x})$ is the image derivative in direction $u$. Wang and Brady [28] characterize the curvature response using the Laplacian and the gradient magnitude,

$$K_{Wang\&Brady}(\mathbf{x}) = (\nabla^2 I(\mathbf{x}))^2 - s|\nabla I(\mathbf{x})|^2.$$
(8)

Kitchen and Rosenfeld [8] present an operator defined as the product between the magnitudes of the gradient and the gradient's change of direction,

$$K_{K\&R}(\mathbf{x}) = \frac{I_{xx}(\mathbf{x})I_y^2(\mathbf{x}) + I_{yy}I_x^2 - 2I_{xy}I_yI_x}{I_x^2 + I_y^2}.$$
(9)

It is obvious that the possible models for detecting interesting features are quite diverse, and still yet there are others, e.g. based on catastrophe theory [19], power law models [3], and simple analysis of local intensity variations designed for fast detection [11].

## 2.2 How to evaluate IP detection

IPs are commonly defined as image pixels that exhibit a high degree of local image variation with respect to a particular local measure. However, this definition is ambiguous regarding how IPs should be interpreted semantically. For instance, it is reasonable to state that IPs must be unique, distinctive, robust and invariant. Nevertheless, all of these properties can be interpreted in different ways and comprehensive measures are not trivially defined. Hence, a simpler task is to list a set of traits that IPs are expected to exhibit and then verify if a given detector can locate points that fulfill those traits. Kenney *et al.* [7], for example, define a set mathematical axioms that can be evaluated analytically. These axioms are based on specific properties of the autocorrelation matrix and how it relates with an idealized image corner. They demonstrate that Shi & Tomasi's detector [23] is the only one that satisfies all four axioms. However, two observations are relevant. First, the axioms assume that an image corner is the only *interesting* image feature, which cannot be true in many real-world scenes. Second, Kenney *et al.* assume that the autocorrelation matrix is essential for IP detection, which is only shown to be true for an optical-flow problem and it is not evident that it will be true in other domains, such as object recognition or image indexing.

Another way that IP detection can be evaluated is to use a set of experimentally testable properties. Schmid *et al.* [22] introduced two such measures: 1) the repeatability rate; and 2) the amount of information conveyed by the local neighborhood around each IP. The former is a measure of invariance with respect to changes in the imaging conditions. While the latter quantifies the amount of dispersion when image information around each IP is mapped to a predefined descriptor space. The experimental comparison conducted by Schmid *et al.* showed that an "*improved*" Harris detector $K_{Harris}$ yields the best overall performance score. However, the experimental setup given in that work has several shortcomings. First, several detectors are left out of the comparison, such as the determinant of the Hessian proposed by Beaudet [2]. Second, Schmid *et al.* estimate the information content based on the local jet at each IP [9]. However, subsequent results by Mikolajczyk and Schmid [14] suggest that the SIFT descriptor [13] gives a more complete and discriminant characterization of image content. Thirdly, even if repeatability and information content are reasonable criteria, other properties could be desirable. For instance, the amount of global separability between detected IPs, which represents the dispersion of IPs within an image [20, 25, 26]. Finally, the evaluation presented by Schmid *et al.* does not study the relationship between each criteria, i.e., whether or not conflicts exist between the objectives and how those conflicts might be accounted for.

Still yet, the choice of a detector is also influenced by the specific requirements of an application. An example can be seen in the work by Rebai *et al.* [20] where IPs are desired to be centered on higher level semantic concepts in order to simplify a subsequent object recognition stage. Li *et al.* [12], for instance, develop a new IP
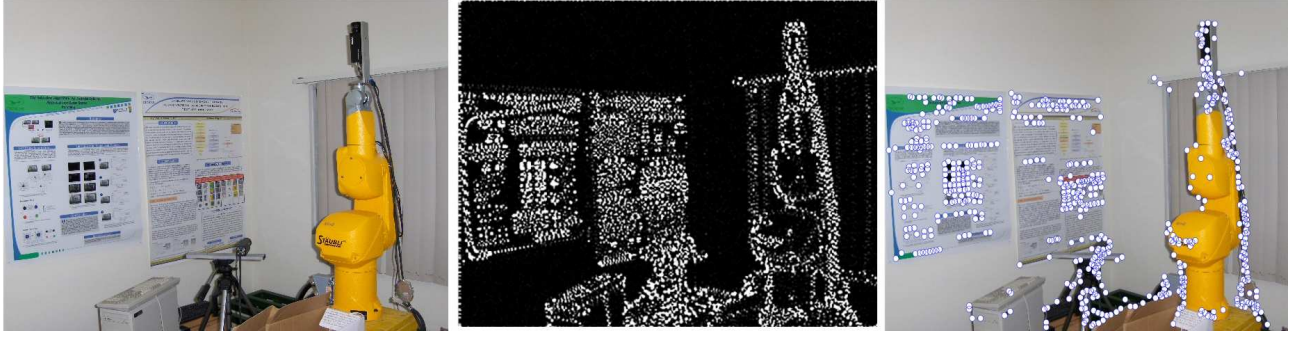
**Figure 1: A look at interest point detection: Left, an input image $I$; Middle, *interest image* $I^*$; Right, detected points after non-maximum suppression and thresholding superimposed on $I$.**

selection scheme that is adapted to improve performance in weakly textured images. Their scheme is tested on a face recognition task with good results. Lepetit *et al.* [11] implement a point detection method designed for a real-time computation used for efficient and robust feature matching. Weijer *et al.* [27] perform color boosting in order to facilitate IP detection on color images. Two final examples are the works by Davison *et al.* [4] and Yang *et al.* [29]. In each of those works, the amount of point dispersion is considered paramount to their application of monocular SLAM and image registration respectively. In both cases, the authors select a previously proposed detector. However, it is reasonable to assume that a feature detector that is optimized considering the special requirements of each problem domain could help to improve the systems' overall performance. Summarizing, because of the large amount of different IP detection methods proposed thus far, choosing a detector can be a non-trivial task. Thence, many researchers have attempted to design detectors that fulfill their application's specific requirements using several criteria.

## 2.3 Proposal

The present work presents a MO-GP approach that can automatically design IP operators. The MO formulation allows for the inclusion of different types of performance criteria within the evaluation process. Additionally, conflicts between objectives are accounted for in a principled manner by defining optimality based on Pareto dominance relations. Thus, the proposed scheme permits the algorithm to generate a set of optimal solutions from which a system designer may choose from. Finally, this paper presents several IP operators which despite their unorthodox sequence of operations are able to yield a higher performance than standard computer vision detectors [22].

## 3. MEASURES OF PERFORMANCE

The following performance criteria are considered by the proposed MO-GP search: 1) the repeatability rate; 2) the amount of global separability; and 3) the information content provided by the set of detected IPs. Additionally, the MO formulation allows the incorporation of other measures in a straightforward manner.

## 3.1 Repeatability rate

The stability is measured through the repeatability rate that estimates how detection is independent of the imaging conditions [22]. An interest point $\mathbf{x}_1$ detected in image $I_1$ is repeated in image $I_i$ if the corresponding point $\mathbf{x}_i$ is detected in image $I_i$. In the case of planar scenes, a relation between points $\mathbf{x}_1$ and $\mathbf{x}_i$ can be established with the homography $H_{1,i}$, where $\mathbf{x}_i = H_{1,i}\mathbf{x}_1$, see
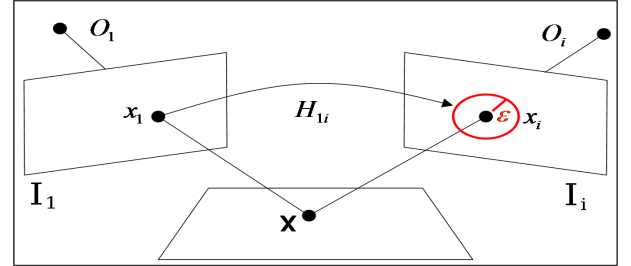


**Figure 2: A 3D point is projected onto points $\mathbf{x}_1$ and $\mathbf{x}_2$ on images $I_1$ and $I_2$ respectively. $x_1$ is said to be repeated by $\mathbf{x}_i$, if a point is detected within a neighborhood of $\mathbf{x}_i$ of size $\epsilon$. For planar scenes $\mathbf{x}_1$ and $\mathbf{x}_i$ are related by the homography $H_{1,i}$.**

Figure 2. The repeatability rate measures the number of repeated points between both images with respect to the total number of detected IPs. A repeated IP is said to be detected at pixel $\mathbf{x}_i$ if it lies within a given neighborhood of size $\epsilon = 1.5\ pixels$. The set of point pairs $(\mathbf{x}_1^c, \mathbf{x}_i^c)$ that lie in the common part of both images and correspond within an error $\epsilon$ is defined by

$$R_{I_i}(\epsilon) = \left\{ (\mathbf{x}_1^c, \mathbf{x}_i^c) \,|\, dist\left(H_{1,i}\mathbf{x}_1^c, \mathbf{x}_i^c\right) < \epsilon \right\}. \qquad (10)$$

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:

$$r_{I_i}(\epsilon) = \frac{|R_i(\epsilon)|}{min(\gamma_1, \gamma_i)}, \qquad (11)$$

where $\gamma_1 = |\{\mathbf{x}_1^c\}|$ and $\gamma_i = |\{\mathbf{x}_i^c\}|$ are the total number of points extracted from image $I_1$ and image $I_i$ respectively.

## 3.2 Information content

Schmid *et al.* defined this measure relative to the likelihood of a local descriptor computed at a given IP [22]. For every detected IP $\mathbf{x}$ a corresponding local image descriptor $\gamma$ is computed. Therefore, if we consider that an IP detector identifies a set $X$ of $n$ interest points, there will be a corresponding set of descriptors $\Gamma$, where $\forall\, \mathbf{x} \in X\ \exists\, \gamma \in \Gamma$. Moreover, let $\Upsilon$ represent the space of all possible descriptors; thus, if the descriptors in $\Gamma$ are crowded within a small region of $\Upsilon$, then the set $X$ conveys a small amount of information content denoted by $\mathcal{I}$, with the converse being true in the opposite case. Therefore, when descriptors are used for image matching problems then the set $\Gamma$ that maximizes the probability $p(I|\Gamma)$ is the set of descriptors that best describes image $I$. Based on information theory, $\mathcal{I}$ is obtained using the amount of entropy

contained within the set of the descriptors $\Gamma$. Therefore, if we consider a partition $\Upsilon = \{\Upsilon_j\}$, and the probability $p_j$ is given by the histogram count of descriptors $\gamma \in \Upsilon_j$ from the set $\Gamma$, then the information content of the set $X$ of detected IPs in $I$ is given by

$$\mathcal{I}(\Gamma) = - \sum p_j \cdot log_2(p_j) \,. \qquad (12)$$

Therefore, it is necessary to select a local descriptor in order to compute $\mathcal{I}$. In the present work the SIFT descriptor is employed because it compares favorably with other methods [14].

### 3.2.1  Scale Invariant Feature Transform: SIFT

The SIFT descriptor is based on the gradient distribution within a detected region and it is computed in the following manner [13]. Assume that an IP is detected at image pixel $\mathbf{x}$; then, an image patch $P$ centered on $\mathbf{x}$ is used to construct the corresponding descriptor $\gamma$ with the size of $P$ $41 \times 41$ pixels in this work. The SIFT descriptor represents a 3D histogram of gradient locations and orientations on $P$, where the contribution to the location and orientation bins is weighted by the gradient magnitude at each pixel. To build the histogram, the location is quantized as a $4 \times 4$ grid within $P$, and the gradient angle is quantized into eight orientation bins centered at $\{0, \frac{\pi}{4}, \frac{\pi}{2}, ..., \frac{7 \cdot \pi}{4}\}$. Therefore, the SIFT descriptor has a total of 128 dimensions. The descriptor is robust to small distortions and localization errors. Additionally, the descriptor is normalized by the square root of the sum of squared components to account for illumination invariance.

In order to compute an entropy measure a 128-dimensional grid is required, something that is computationally infeasible. To simplify the computation process each SIFT dimension is considered to be independent and a mean entropy value $\mathcal{I}_\mu$ of all 128 dimensions is used to estimate an operator's information content. SIFT values are normalized between $[0, 1]$ and each dimension is divided into 40 bins to compute the entropy value.

## 3.3  Global separability

It is also appropriate to use an entropy based measure for the amount of point dispersion. In this case, the entropy is computed from the partition $I = \{I_j\}$ of the spatial dimensions of the image plane. Hence, $\mathcal{D}$ is the entropy value of the spatial distribution of detected interest points within the image, $\mathcal{D}(I, X) = -\sum P_j \cdot log_2(P_j)$, where $P_j$ is approximated by the 2D histogram of the position of IPs within $I$. The image is divided into a 2D grid where each bin has a size of $8 \times 8$ pixels. Because point dispersion depends on the manner in which non-maximum suppression is carried out, a window size of $5 \times 5$ was used for every detector discussed here.

## 4.  MULTIOBJECTIVE OPTIMIZATION

In this work, IP detection is studied in MO terms because of the possible conflicts between the previously defined performance criteria. Multiobjective optimization is a principled approach to solving problems where objectives present conflicts between them. When a MO problem lacks a closed form solution, it is necessary to rely on computational search methods in order to obtain an approximation for the true Pareto-Optimal Set. One approach is to employ MO evolutionary algorithms (MOEAs) which are capable of performing parallel and distributed search. Nowadays, state-of-the-art MOEAs are expected to converge towards representative sampling of the true Pareto Front. However, real-world problems present serious challenges for these search algorithms, such as non-linear and disconnected objective spaces, constraint satisfaction, isolated minima, and combinatorial aspects, to name but a few. These and
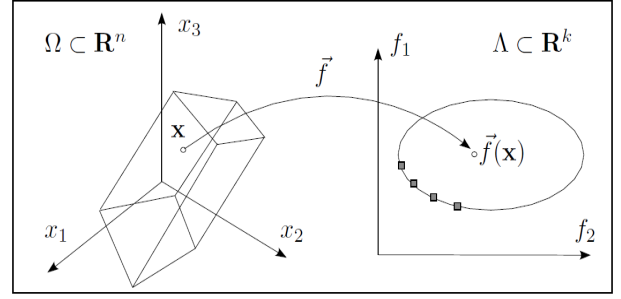


**Figure 3: *Decision* and *Objective* spaces for MO optimization. A solution parametrization $\mathbf{x}$ is mapped by a vector function $\vec{f}$ into a vector in objective function space. The highlighted points on the boundary of $\Lambda$ are elements of the Pareto Front.**

other considerations makes MO problem solving a non-trivial task when it is applied in real-world situations.

In MO optimization it is necessary to consider two different and complimentary spaces: one for decision variables and another for the objective functions, see Figure 3. In the case of real valued functions, these two spaces are related by the mapping $\vec{f} : \mathcal{R}^n \rightarrow \mathcal{R}^k$. Constraints on the objective vector $\vec{f}(\mathbf{x}) = [f_1(x), ..., f_k(x)]$ define a feasible region $\Omega \subset \mathcal{R}^n$ in the decision space along with its corresponding image $\Lambda \subset \mathcal{R}^n$ on the objective function space. Now, the following concepts define the concept of optimality in a MO optimization problem [1]. The optimum is thereby found at the frontier of the objective space called the *Pareto Front*, while the values of the corresponding decision variables in $\Omega$ are called the *Pareto-Optimal Set*. The optimal solutions satisfy the non-dominance relations as defined below [18].

*Definition 2.* **Pareto dominance**: Given $k$ objectives and an ordered set $N = \{1, ..., k\}$, an objective vector $\vec{f}^u$ is said to dominate another objective vector $\vec{f}^v$ (written as $\vec{f}^u \preceq \vec{f}^v$) $\Leftrightarrow \forall\, i \in N$, $f_i^u \leq f_i^v \,\wedge\, \exists\, j \in N \mid f_j^u < f_j^v$.

*Definition 3.* **Pareto optimality:** A solution vector $\mathbf{x}^* \in \Omega$ is optimal if $\forall\, \mathbf{x} \in \Omega$ it is true that $\forall\, i \in N$, $f_i(\mathbf{x}^*) = f_i(\mathbf{x}) \vee \exists\, i \in N \mid f_i(\mathbf{x}^*) < f_i(\mathbf{x})$.

*Definition 4.* **Pareto-Optimal Set:** For a multiobjective problem $\vec{f}(\mathbf{x})$, the set of Pareto optimal solutions is given by the set $\mathcal{P}^* = \left\{ \mathbf{x} \in \Omega \mid \nexists\, \mathbf{x}' \in \Omega \; that \; \vec{f}(\mathbf{x}') \preceq \vec{f}(\mathbf{x}) \right\}$.

*Definition 5.* **Pareto Front:** For a multiobjective problem with an objective vector $\vec{f}(\mathbf{x})$ with a pareto optimal set $\mathcal{P}^*$, the Pareto Front is defined as $\mathcal{PF}^* = \{ \mathbf{u} = (f_1(\mathbf{x}), ..., f_k(\mathbf{x})) | \mathbf{x} \in \mathcal{P}^* \}$.

## 5.  SYNTHESIS OF IP OPERATORS

### 5.1  Previous work

Previous work by Trujillo & Olague proposed an automatic design process for IP operators using GP [25, 26]. In those works, however, a mono-objective problem was formulated by combining the repeatability score and the amount of global separability into a single objective function. Using two solutions as examples, Trujillo & Olague were able to confirm that the GP-based design is capable of synthesizing reliable and competitive operators. One of

---

[1]A minimization problems is described without loss of generality.

the operators performed a simple Difference-of-Gaussian filtering (DoG), while the other was a modified version of the Hessian operation proposed by Beaudet.

$$K_{IPGP1}(\mathbf{x}) = G_{\sigma=2} * (G_{\sigma=1} * I - I) , \qquad (13)$$

$$K_{IPGP2}(\mathbf{x}) = G_{\sigma=1} * [L_{xx}(\mathbf{x}) \cdot L_{yy}(\mathbf{x}) - L_{xy}^2(\mathbf{x})] . \qquad (14)$$

In the latter case, the Hessian-based operator provides a similar improvement to the one proposed by Schmid *et al.* for the Harris detector [22]. However, the previous proposal by Trujillo & Olague does not account for the interdependencies or possible conflicts between the different objectives employed as part of the evaluation process. Moreover, it is not evident how other objectives can be added to the mono-objective formulation proposed in that earlier work. These considerations reveal that a more comprehensive formulation is required, which is addressed in this paper.

## 5.2 Multiobjective approach

The design of IP operators using a GP process is proposed as follows. The function set $F$ and the terminal set $T$ are defined in such a way that it is conceivable that the GP search process can build any of the operators $K$ discussed in Section 2.1 as well as other, possibly novel, designs.

$$F = \{+, |+|, -, |-|, |I_{out}|, *, \div, I_{out}^2, \sqrt{I}, log_2(I_{out}),$$
$$k \cdot I_{out}, \tfrac{\delta}{\delta x} G_{\sigma_D}, \tfrac{\delta}{\delta y} G_{\sigma_D}, G_{\sigma=1}, G_{\sigma=2}\} ,$$
$$T = \{I, L_x, L_{xx}, L_{xy}, L_{yy}, L_y\} ,$$

where $I$ is the input image, and $I_{out}$ is any of the terminals in $T$ or the output of any of the functions in $F$, $L_u$ is the image derivative along direction $u$, $G_u$ are Gaussian smoothing filters, $\tfrac{\delta}{\delta u} G_{\sigma_D}$ the derivative of a Gaussian function[2], and the constant $k = 0.05$. Individual solutions are evaluated using the following cost functions defined for minimization:

- Global separability: $f_1(K) = \dfrac{1}{\exp(\mathcal{D}(I, X) - c_1)}$ .

- Information content: $f_2(K) = \dfrac{1}{\exp(\mathcal{I}_\mu(\Gamma) - c_2)}$ .

- Stability: $f_3(K) = \dfrac{1}{r_{K,J}(\epsilon) + \phi}$ .

The constants were set experimentally to $\epsilon = 0.001$, $c_1 = 10$ and $c_2 = 2.8$. Because these measures are based on experimental performance, a training set is required to compute a performance score. Therefore, $r_{K,J}(\epsilon)$ is the average repeatability rate of an operator $K$ computed on a training sequence $J$. On the other hand, $D(I, X)$ and $I_\mu(\Gamma)$ are computed using only the base image of $J$. Previous results by Trujillo & Olague [26] suggest that using only one image sequence during the evaluation process is enough to produce robust and general operators. In the proposed MO approach, the selection and survival strategy must account for Pareto dominance relations. Therefore, the SPEA2 algorithm, a third generation MOEA, is used to perform population management [30]. SPEA2 uses a fitness assignment that accounts for both domination and non- domination relations between individuals in both the current population and the population archive. Diversity preservation is achieved by using a k-*th* nearest neighbor clustering algorithm that penalizes individuals that reside in densely populated regions of objective space. Additionally, the algorithm uses a fixed-size archiving approach and a truncation scheme that promotes diversity by removing individuals that have the minimum distance to

---

[2]All Gaussian filters are applied by convolution.

its neighbors. Finally, it preserves boundary solutions by using a carefully designed selection operator.

## 5.3 A set of testable hypothesis

This work considers multiple criteria concurrently within the optimization process and deals with conflicts between these criteria comprehensively. In order to experimentally study these conflicts and interdependencies between the proposed objectives, the following hypothesis are proposed:

1. **Hypothesis A:** The criteria of *information content* and *point dispersion do not* represent conflicting objectives for an *average image* (referred to as *HA* hereafter).

2. **Hypothesis B:** The properties of *global separability* and *repeatability* represent conflicting objectives for an *average image* (*HB*).

3. **Hypothesis C:** The properties of *information content* and *repeatability* represent conflicting objectives for an *average image* (*HC*).

Each of these hypothesis was formulated in order to experimentally test if conflicts can be confirmed or denied between each pair of IP performance measures. It is assumed that if a Pareto Front is observed then this is taken as positive evidence that a single non-dominated solution does not exist. In all three cases a reference is made to the idea of an *average image*, which is intended to refer to *most images of real-world scenes*, which is conceptually useful. Each hypothesis is intuitive, for instance *HA* suggests a strong correlation between global separability and information content, because most scenes do not contain a repetitive structure that covers all of the image. In other words, the local regions around most IPs are expected to be different if detected IPs are spread out across the image. On the other hand, in the case of *HB* it is simple to conceive a degenerative detector that identifies IPs crowded together on isolated portions of the image that manages a high repeatability rate. Finally, *HC* is based on the assumption that *HA* is true from which a similar argument to the one described for *HB* can be derived. In sum, falsifying *HA*, *HB* and *HC* could be done experimentally using the MO-GP implementation described here.

## 6. EXPERIMENTATION AND RESULTS

The algorithm was programmed using the Matlab toolbox GPLAB[3]. The Matlab code was coupled with an implementation of the SPEA2 selection algorithm, written in C++, provided by the *Platform and Programming Language Independent Interface for Search Algorithms*[4]. The image sequence used for training was the Van Gogh set of a planar scene with rotation transformations provided by the Visual Geometry Group at Oxford University, along with Matlab source code to compute the repeatability rate[5]. The control parameters for the MO/GP search are presented in Table 1. The MO search was carried out four times for each experimental configuration; in each case a different maximum depth was allowed for the corresponding program trees. The following subsections describe the experiments conducted for each of the three hypothesis: *HA*, *HB* and *HC*. In each case, three operators are presented and sample IPs are shown on the training image. Additionally, the Pareto Front found in each experiment is shown. In order to verify the generality of the measures for global separability and information

---

[3]http://gplab.sourceforge.net/index.html, GPLAB A Genetic Programming Toolbox for MATLAB by Sara Silva
[4]http://www.tik.ee.ethz.ch/sop/pisa/
[5]http://www.robots.ox.ac.uk/ vgg/research/

content, the Pareto dominance relations were tested on eight additional test images shown in Figure 4. The Pareto-Optimal set of solutions showed a consistent performance on these tests.



**Figure 4: Images used to test the evolved operators.**



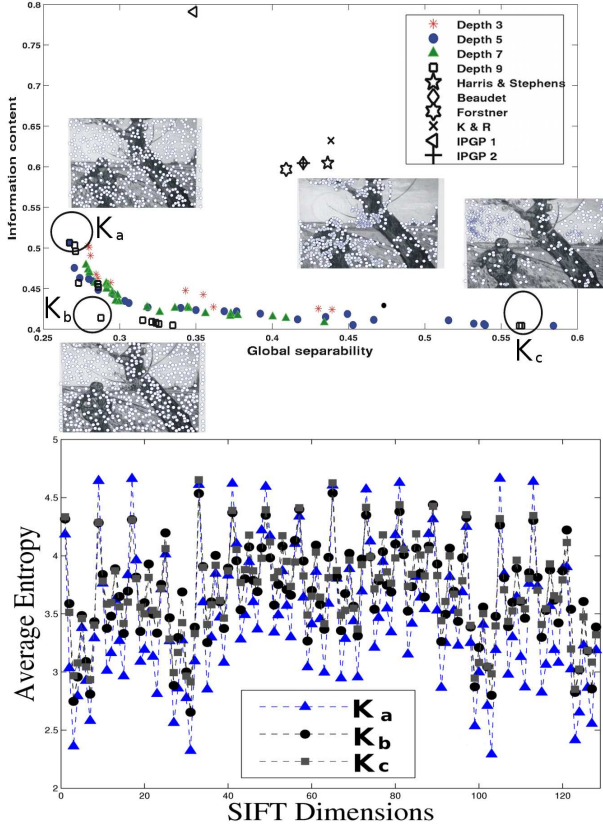**Figure 6: Experimental results of HB. Pareto front and comparison with previous detectors.**



**Figure 5: Experimental results of HA. Top: Pareto front and comparison with previous detectors. Bottom: Entropy computed in each SIFT dimension for operators $K_a$, $K_b$ and $K_c$.**



**Figure 7: Experimental results of HC. Top: Pareto front and comparison with previous detectors. Bottom: Entropy computed in each SIFT dimension for operators $K_u$, $K_v$ and $K_w$.**

## 6.1 Experiments: HA

In this configuration the MO optimization focuses on $f_1$ and $f_2$. Figure 5 presents the outcome of the experimental runs. Three inflection points can be seen in the Pareto Front, denoted as $K_a$, $K_b$ and $K_c$; see Table 2. Also shown are the performance of several IP operators, including those evolved using a mono-objective approach by Trujillo & Olague: $IPGP1$ and $IPGP2$ [25, 26]. Other operators included for comparison are those proposed by Beaudet [2], Harris & Stephens [6], Förstner [5], and Kitchen & Rosenfeld [8]. IPs detected with each operator are also shown, as well as those detected with the Harris & Stephens approach.
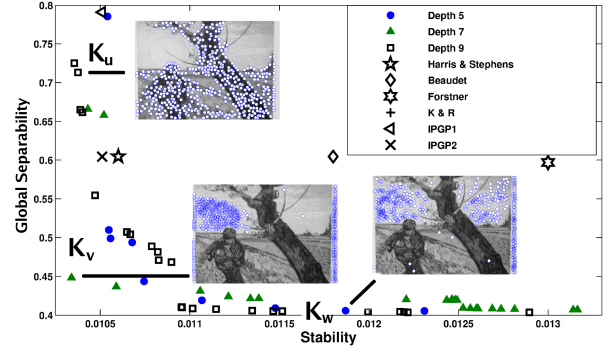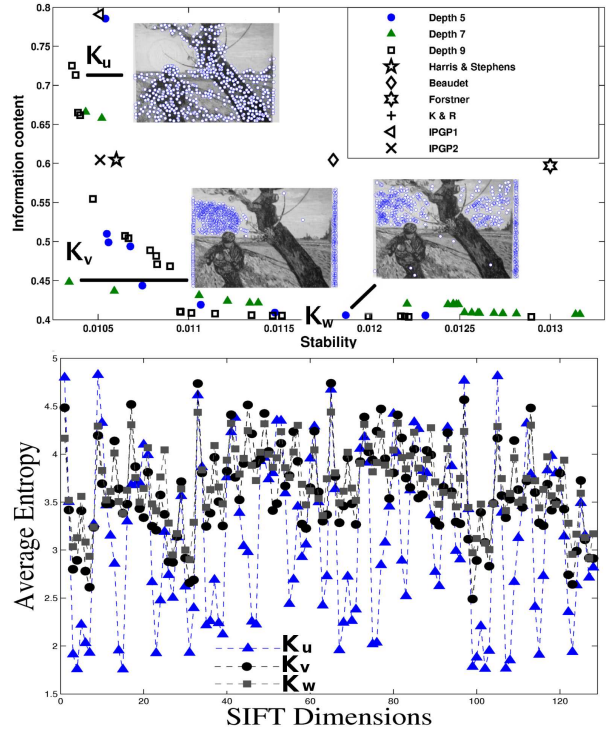
The first obvious assertion is that all previous operators are dominated by the Pareto-Optimal set of solutions that were found. This could be expected from the evolved operators of Trujillo & Olague that do not explicitly account for information content; however, the Harris & Stephens detector performs worse than anticipated. The plots of Figure 5 suggest that *HA* is false. However, making such an assertion has to be done with caution because of two observations. First, it is obvious that $f_2(K_a) > f_1(K_b)$. However, the difference in global separability could be dismissed because $f_1(K_a) \approx f_1(K_b)$. Second, a comparison between $K_b$ and $K_c$ shows that $f_1(K_c) > f_1(K_b)$, and a Kolmogorov-Smirnov test between each SIFT dimension of $\Gamma_b$ and $\Gamma_c$ suggests that regarding $f_2(K_b) \approx f_2(K_c)$ could also be judicious. Therefore, if the Pareto dominance relations are analyzed under these considerations then

**Table 1: General parameter settings for the MO/GP search.**

| Parameters | Description and values |
|---|---|
| *Population size* | 200. |
| *Iterations* | 50. |
| *Initialization* | Ramped Half-and-Half. |
| *Crossover & Mutation prob.* | Crossover prob. $p_c = 0.85$; mutation prob. $p_\mu = 0.15$. |
| *Max depth* | 3,5,7 9 levels. |
| *Archive size* | The SPEA2 archive size: 100. |
| *Selection size* | The amount of individuals selected by SPEA2: 100. |

$K_b$ could be seen as a solution that dominates all other solutions found. Hence, the authors conclude that not enough evidence exists to reject *HA*.

### 6.1.1 Discussion

The results related to information content are unexpected. The operator $K_a$, which is inversely proportional to the curvature along one of the principal directions, transmits the least amount of information. Even more unexpected is the fact that operator $K_c$ conveys the most information content even if the IPs detected are mostly crowded together. Therefore, Figure 5 shows the entropy in each SIFT dimension computed for each $\Gamma_a$, $\Gamma_b$ and $\Gamma_c$. This plot shows that for $K_a$, several dimensions are poorly distributed, while at the same time others have a high amount of dispersion. These results suggest an interesting possibility because the descriptors in $\Gamma_a$ are all very similar in so many SIFT dimensions. Hence, it may be possible to extract a representative global-SIFT descriptor for an entire image. Experiments using the test images in Figure 4 show that every image responds similarly to this operation.

## 6.2 Experiments: HB

The experimental configuration is similar to the previous one. In this case using the fitness measures $f_1$ and $f_3$, see Figure 6. Indeed, the approximation of the Pareto Front for this experiment shows that these objectives are in conflict. Furthermore, the results show some interesting trends from the previously proposed IP operators. First, most designs perform quite poorly concerning the amount of IP dispersion. For example, the Harris & Stephens detector approaches the Pareto Front in the Stability dimension, but fails considerably in the amount of global separability. This illustrates that human designers are biased to a certain kind of problem modeling. Normally, researchers focus mainly on detecting a specific and easily definable image feature, e.g. corners. Second, the operators $IPGP1$ and $IPGP2$ that were proposed by Trujillo & Olague dominate all other operators and $IPGP1$ in particular lies on the approximated Pareto Front. Thus, it is evident that these operators benefited from the fitness measure that incorporated the amount of *stability* and *global separability* explicitly.

## 6.3 Experiments: HC

These experiments also produced a Pareto Front that exhibits a conflict between the objectives $f_2$ and $f_3$. Moreover, the four man-made operators used for comparison differ only in the amount of stability while maintaining very similar information content. On the other hand, each of the evolved operators achieve a different performance. $IPGP2$ is similar, and slightly dominates the Harris & Stephens operator, as well as the other man-made designs. While $IPGP1$ has a low information content and exhibits a large degree of stability, this gives it an extreme position on the Pareto Front presented in Figure 7. Also shown in the figure are the entropy comparisons between three different operators, $K_u$, $K_v$ and $K_w$.

**Table 2: Some of the operators found with the MO/GP search.**

**Experiment: HA**

$$K_a = k^2 \cdot G_2 * G_2 * \left(\frac{L_y}{L_{yy}}\right)$$

$$K_b = k \cdot G_2 * G_2 * \sqrt{\frac{log(G_1 * L)}{G_1 * |L_x + L_{yy}|}}$$

$$K_c = \frac{k \cdot G_1 * I}{L_{yy}^2 \cdot |L_{yy}L_{xx} + \sqrt{L_{xx}}| \cdot |G_1 * I + |G_1 * L_{xy}^2 - L_{yy}||}$$
$$- \left| log\left(G_1 * \frac{L_{xy}}{||L_y - L_{xx}| - |L_x + L_{yy}||}\right) - \frac{G_1 * L_{yy}}{|L_{xx}|}\right|$$

**Experiment: HB**

$$K_p = G_2 * \Big|G_1 * log(G_1 * I^2) +$$
$$G_2 * (G_1 * I - I) + \left|\frac{G_1 * I}{I}\right| \Big|^2$$
$$K_q = G_2 * |G_1 * log(G_1 * I^2) +$$
$$k \cdot G_2 * |G_1 * I - I| + \left|\frac{G_1 * I}{I}\right| |^2$$
$$K_r = G_2 * \left(\frac{L_y}{L_{yy}}\right)$$

**Experiment: HC**

$$K_u = \frac{G_1 * G_1 * G_2 * |G_1 * I|}{G_2 * |G_2 * G_2 * I|}$$
$$K_v = ||L_{yy} + G_1 * G_1 * G_1 * I| + G_1 * L_{xx}|$$
$$K_w = \frac{G_2 * G_1 * I}{(L_{yy} \cdot G_2 * L_y) \cdot (G_2 * L_{xx}) \cdot (k \cdot G_1 * L_{xx})}$$

The comparison shows that even if the points detected by $K_u$ are more sparsely distributed accross the image, its corresponding entropy value in most SIFT dimensions is far lower than the obtained with $K_v$ and $K_w$.

## 7. SUMMARY AND CONCLUSIONS

In this work a Multiobjective Genetic Programming algorithm was described that synthesizes several IP operators using a single program execution. The function and terminal sets rely on common image operations used as primitives that allow the search process to synthesize high-performance operators. Three well-established performance criteria are employed during fitness evaluation. These measures have been identified by the computer vision community as useful in evaluating the performance of IP detectors, these are: the repeatability rate, the amount of information content, and the level of global separability of the detected IPs. In order to test the MO problem formulation three hypothesis were given. These hypothesis were proposed in order to determine wether conflicts exist between the defined performance criteria. The experimental runs suggest that conflicts do indeed exist, and that previously proposed operators do not provide optimal solutions based on the Pareto criterion for optimality. Therefore, based on the experimental evidence it is concluded that IP detection constitutes an ill-posed multiobjective problem. Furthermore, the GP search, using the SPEA2 selection scheme, was able to generate a diverse set of Pareto optimal solutions, and several examples were presented that outperform man-made designs. Future work could focus on incorporating other performance measures that are dictated by the requirements of a specific application, this is possible because the proposed MO scheme can be extended in a principled manner.

# 8. REFERENCES

[1] H. Asada and M. Brady. The curvature primal sketch. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(1):2–14, 1986.

[2] P. R. Beaudet. Rotational invariant image operators. In *Proceedings of the 4th International Joint Conference on Pattern Recognition (ICPR 1978),Tokyo, Japan*, pages 579–583, 1978.

[3] Y. Caron, P. Makris, and N. Vincent. Use of power law models in detecting region of interest. *Pattern Recogn.*, 40(9):2521–2529, 2007.

[4] A. J. Davison and N. D. Molton. Monoslam: Real-time single camera slam. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(6):1052–1067, 2007. Member-Ian D. Reid and Member-Olivier Stasse.

[5] W. Förstner and E. Gülch. 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, 1987.

[6] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings from the Fourth Alvey Vision Conference*, volume 15, pages 147–151, 1988.

[7] C. S. Kenney, M. Zuliani, and B. S. Manjunath. An axiomatic approach to corner detection. In *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1*, pages 191–197, Washington, DC, USA, 2005. IEEE Computer Society.

[8] L. Kitchen and A. Rosenfeld. Gray-level corner detection. *Pattern Recognition Letters*, 1:95–102, December 1982.

[9] J. J. Koenderink and A. J. van Doom. Representation of local geometry in the visual system. *Biol. Cybern.*, 55(6):367–375, 1987.

[10] I. Laptev and T. Lindeberg. Space-time interest points. In *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision*, page 432, Washington, DC, USA, 2003. IEEE Computer Society.

[11] V. Lepetit and P. Fua. Keypoint recognition using randomized trees. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(9):1465–1479, 2006.

[12] Q. Li, J. Ye, and C. Kambhamettu. Interest point detection using imbalance oriented selection. *Pattern Recogn.*, 41(2):672–688, 2008.

[13] D. G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the International Conference on Computer Vision, 20-25 September, 1999, Kerkyra, Corfu, Greece*, volume 2, pages 1150–1157. IEEE Computer Society, 1999.

[14] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005.

[15] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool. A comparison of affine region detectors. *International Journal of Computer Vision*, 65(1-2):43–72, 2005.

[16] H. P. Moravec. Towards automatic visual obstacle avoidance. In *IJCAI*, page 584, 1977.

[17] G. Olague and B. Hernández. A new accurate and flexible model based multi-corner detector for measurement and recognition. *Pattern Recognition Letters*, 26(1):27–41, 2005.

[18] V. Pareto. *Cours D'Economie Politique*. Rouge, Lausanne, 1896.

[19] B. Platel, E. Balmachnova, L. Florack, and B. M. ter Haar Romeny. Top-points as interest points for image matching. In A. Leonardis et al., editors, *Proceedings from ECCV 2006, 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006, Part I*, volume 3951 of *Lecture Notes in Computer Science*, pages 418–429. Springer, 2006.

[20] A. Rebai, A. Joly, and N. Boujemaa. Interpretability based interest points detection. In *CIVR '07: Proceedings of the 6th ACM international conference on Image and video retrieval*, pages 33–40, New York, NY, USA, 2007. ACM.

[21] K. Rohr. Recognizing corners by fitting parametric models. *Int. J. Comput. Vision*, 9(3):213–230, 1992.

[22] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of interest point detectors. *International Journal of Computer Vision*, 37(2):151–172, 2000.

[23] J. Shi and C. Tomasi. Good features to track. In *Proceedings of the 1994 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94), June 1994, Seattle, WA, USA*, pages 593–600. IEEE Computer Society, 1994.

[24] P. Tissainayagam and D. Suter. Assessing the performance of corner detectors for point feature tracking applications. *Image Vision Comput.*, 22(8):663–679, 2004.

[25] L. Trujillo and G. Olague. Synthesis of interest point detectors through genetic programming. In M. Cattolico, editor, *Proceedings of the Genetic and Evolutionary Computation Conference, GECCO 2006, Seattle, Washington, USA, July 8-12, 2006*, volume 1, pages 887–894. ACM, 2006.

[26] L. Trujillo and G. Olague. Using evolution to learn how to perform interest point detection. In *Proceedings of the 18th International Conference on Pattern Recognition (ICPR 2006), 20-24 August 2006, Hong Kong, China*, volume 1, pages 211–214. IEEE Computer Society, 2006.

[27] J. van de Weijer, T. Gevers, and A. D. Bagdanov. Boosting color saliency in image feature detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(1):150–156, 2006.

[28] H. Wang and J. Brady. Corner detection for 3d vision using array processors. In *Proceedings from BARNAIMAGE 91, Barcelona, Spain*, Secaucus, NJ, USA, 1991. Springer-Verlag.

[29] G. Yang, C. V. Stewart, M. Sofka, and C.-L. Tsai. Registration of challenging image pairs: Initialization, estimation, and decision. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(11):1973–1989, 2007.

[30] E. Zitzler, M. Laumanns, and L. Thiele. Spea2: Improving the strength pareto evolutionary algorithm for multiobjective optimization. In *Evolutionary Methods for Design, Optimisation, and Control*, pages 19–26, 2002.