Regularity based descriptor computed from local image oscillations

Leonardo Trujillo[a], **Gustavo Olague**[a*], **Pierrick Legrand**[a,b,c] **and Evelyne Lutton**[b]

[a]*EvoVisión Project, CICESE Research Center, Applied Physics Division, Km. 107 carretera Tijuana-Ensenada 22860, Ensenada, B.C. México*

[b]*Complex Team, INRIA Roquencourt, Domaine de Voluceau, BP 105, 78153 Le Chesnay Cedex, France*

[c]*U.F.R Sciences et Modélisation, Université de Bordeaux, Laboratoire de Mathématiques appliquées, UMR CNRS 5466, France*

{olague, trujillo}@cicese.mx

Abstract: This work presents a novel local image descriptor based on the concept of pointwise signal regularity. Local image regions are extracted using either an interest point or an interest region detector, and discriminative feature vectors are constructed by uniformly sampling the pointwise Hölderian regularity around each region center. Regularity estimation is performed using local image oscillations, the most straightforward method directly derived from the definition of the Hölder exponent. Furthermore, estimating the Hölder exponent in this manner has proven to be superior, in most cases, when compared to wavelet based estimation as was shown in previous work. Our detector shows invariance to illumination change, JPEG compression, image rotation and scale change. Results show that the proposed descriptor is stable with respect to variations in imaging conditions, and reliable performance metrics prove it to be comparable and in some instances better than SIFT, the state-of-the-art in local descriptors.

OCIS codes: (110.2960) Image analysis; (100.5010) Pattern recognition and feature extraction; (100.5760) Rotation-invariant pattern recognition.

References and links

1. H. P. Moravec, "Towards automatic visual obstacle avoidance," in *IJCAI*, pp. 584 (1977).
2. L. Trujillo and G. Olague, "Synthesis of interest point detectors through genetic programming, " in *Proceedings from GECCO 2006*, Mike Cattolico, eds., Vol.1, (ACM Press 2006), pp. 887–894.
3. L. Trujillo and G. Olague, "Using Evolution to Learn How to Perform Interest Point Detection," in *Proceedings from ICPR 2006*, X.Y. Tang *et al.*, eds., Vol. 1, (IEEE 2006), pp. 211–214.
4. D.G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," Int. J. Comput. Vision, **2**, 91–110 (2004).
5. C. Schmid and K. Mikolajczyk, "A Performance Evaluation of Local Descriptors," IEEE Trans. on Patt. Ana. and Mach. Int., **27**, 1615–1630 (2005).
6. K. Falconer, *Fractal geometry, Mathematical Foundations and Applications* (Wiley, 1990).
7. J. Lévy Véhel, "Fractal Approaches in Signal Processing," Fractals, **3**, 755–775 (1995).
8. P. Legrand and J. Lévy Véhel, "Local regularity-based interpolation," in *WAVELET X*, Part of SPIE's Symposium on Optical Science and Technology, SPIE **5207** (2003).
9. P. Legrand, "Debruitage et interpolation par analyse de la regularite Hölderienne. Application a la modelisation du frottement pneumatique-chaussee," P.h.D. thesis, Université de Nantes (2004).

10. C. Tricot, *Curves and Fractal Dimension* (Springer-Verlag 1995).
11. K. Mikolajczyk and C. Schmid, “Scale and affine invariant interest point detectors,” Int. J. Comput. Vision, **1**, 63–86 (2004).
12. Visual Geometry Group: http://www.robots.ox.ac.uk/ vgg/research/

1. Introduction

The feature extraction problem, in the domain of image analysis systems, poses two main research questions. How can *distinctive* areas within an image be **detected**? And, how can distinctive areas be **described** in such a way as to facilitate their identification? The main concepts to be taken from those questions are: detection and description. Concerning the former, the detection problem, a mainstay in vision systems are interest point or interest region extraction algorithms. These techniques search for image pixels, or image regions, that exhibit high signal variations with respect to a particular local measure. Solutions have been designed based on studying intrinsic properties of 2D-signals [1], and more recently a novel methodology has been proposed that solves a properly framed optimization problem and automatically synthesizes interest point operators with Genetic Programming [2, 3]. In response to the second question, dealing with the concept of description, different techniques have been proposed that encode the information within these so called interesting regions. Hence, discriminative feature are constructed that uniquely characterize each interest regions. This in turn allows for efficient feature matching in a wide range of imaging problems. Currently, the SIFT [4] descriptor has proven to be the most discriminative local descriptor in machine vision literature, and shows the highest performance with respect to the current set of benchmark tests [5].

This paper presents a novel region descriptor based on the concept of Hölderian regularity. By approximating the pointwise Hölder exponent, also known as the Lipschitz exponent, using local signal oscillations around each image point, we are able to construct discriminative feature vectors. Our proposed descriptor is invariant to several types of changes in viewing conditions, exhibiting high and stable performance. As such, the main contribution of this work is that it introduces novel concepts to the field of feature extraction algorithms, using formal mathematical tools and corroborated by high performance on standard tests.

The remainder of this paper is organized as follows. Section 2 gives a brief overview of related work. Section 3 presents the concept of Hölderian regularity and how to estimate it. Section 4 introduces our local descriptor based on pointwise Hölder exponents. Later in Section 5, experimental results are provided. Section 6 presents conclusions and outline future work.

2. Related work

It is not our intention to give a comprehensive summary on the subject of local descriptors, such a discussion can be found in Ref. [5]. Hence, we will only focus on presenting the basic strategies followed by the most common type of region detectors, distribution based descriptors, and discuss the SIFT strategy.

Currently, most state-of-the-art local descriptors use a distribution based approach. These techniques characterize image information using local histograms of a particular measure related to shape or appearance. The most simple would be using histograms of pixel intensities, while more complex representations could be based on representing texture characteristics. The most successful descriptor currently available in computer vision literature is SIFT, developed by David Lowe in [4], which builds an histogram of gradient distributions within an interest region. The descriptor builds a 3D histogram of gradient locations and orientations, weighted by the gradient magnitudes. Although SIFT combines both a scale invariant detector with the gradient distribution descriptor, only the latter has proven to outperform other types of techniques, and it is possible to replace the former with a more reliable region detector.

3. Hölder regularity

One of the most popular ways to measure the regularity of a signal, be it pointwise or local, is to consider Hölder spaces. Hence, we will present the concept of regularity expressed through the Hölder exponent.

DEFINITION 1. Let $f : \Re \rightarrow \Re$, $s \in \Re^{+*} \setminus \mathbf{N}$ and $x_0 \in \Re$. Then, $f \in C^s(x_0) \Leftrightarrow \exists \eta \in \Re^{+*}$, a polynom P of degree $< s$ and a constant c such that

$$\forall x \in B(x_0, \eta), |f(x) - P(x - x_0)| \leq c|x - x_0|^s. \tag{1}$$

The pointwise Hölder exponent of f at x_0 is $\alpha_p = \sup_s \{f \in C^s(x_0)\}$ (see Figure 1).

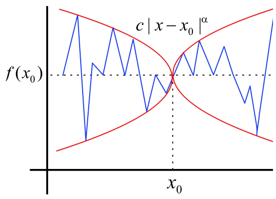

Fig. 1. Hölderian envelope of signal f at point x_0.

The concept of signal regularity, characterized by the Hölder exponent, has been widely used in fractal analysis [6]. With regards to image analysis, the Hölder exponent provides a great deal of information related to the local structure around each point. Hence, it has been applied to such tasks as edge detection [7] and image interpolation [8]. Furthermore, because most local image descriptors are fundamentally attempting to describe local image variations and overall structure, it is a natural conclusion to expect that Hölderian regularity will prove to be a useful tool in this task.

3.1. Estimating the Hölder exponent with oscillations

The most natural way to estimate the Hölder exponent, because it follows from its definition, consists in studying the oscillations around each point. This method gives accurate results, better than those obtained using wavelet analysis in most cases [9], hence it will be the technique of choice to compute our proposed descriptor. A brief description of this technique will now be given, for a more detailed analysis please see Ref. [10]. It is pointed out that the Hölder exponent of function $f(t)$ at t is $\alpha_p \in [0,1]$, if a constant c exists such that $\forall\, t'$ in a vicinity of t,

$$|f(t) - f(t')| \leq c|t - t'|^{\alpha_p}. \tag{2}$$

In terms of signal oscillations, this condition can be written as: a function $f(t)$ is Hölderian with exponent $\alpha_p \in [0,1]$ at t if $\exists c\ \forall \tau$ such that $osc_\tau(t) \leq c\tau^{\alpha_p}$, with

$$osc_\tau(t) = \sup_{|t-t'|\leq\tau} f(t') - \inf_{|t-t'|\leq\tau} f(t') = \sup_{t',t''\in[t-\tau,t+\tau]} |f(t') - f(t'')|. \tag{3}$$

An estimation of the regularity will be built at each point by computing the slope of the regression between the logarithm of the oscillation and the logarithm of the dimension of the neighborhood at which one calculates the oscillation. From an algorithmic point of view, it is preferable not to use all sizes of neighborhoods between two values τ_{min} and τ_{max}. Hence, we calculate the oscillation at point t only on intervals of the form $[t - \tau_r : t + \tau_r]$, where $\tau_r = base^r$. Here, we use least squares regression, with $base = 2$ and $r = 1, 2, \ldots, 7$. For a 2D signal, t defines a point in 2D space and τ_r a radius around t, such that the Euclidian distances $d(t', t)$

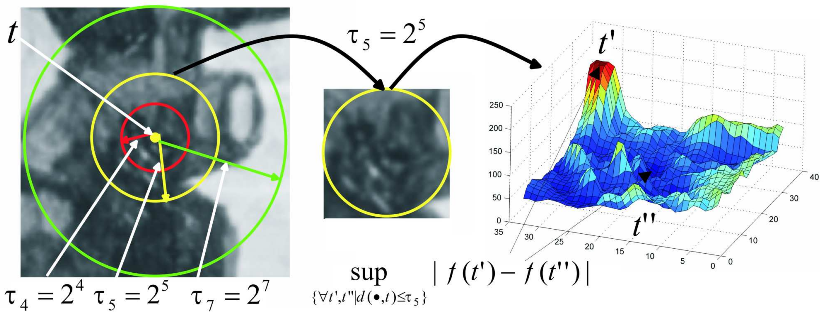

Fig. 2. Estimating the Hölder exponent with oscillations. **Left**: the region of interest λ, and three of the seven neighborhoods around point t, when $r = 1, 2, \cdots, 7$. **Center**: the neighborhood of radius $\tau_5 = 32$ pixels, with *base* $= 2$. **Right**: computing the supremum of the differences within radius τ_5, where d denotes the Euclidian distance.

and $d(t'', t)$ are $\leq \tau_r$. We can visualize this process in Figure 2. The method is reliable under three conditions: that $\alpha_p < 1$, the regression converges, and it converges towards a valid slope.

4. Hölder descriptor

Now that we have described a method to accurately characterize the pointwise signal regularity, we can now move on to describe how we use this information to build our local descriptors. The process, described in Figure 3, is as follows. First, a set Λ of regions of interest are extracted

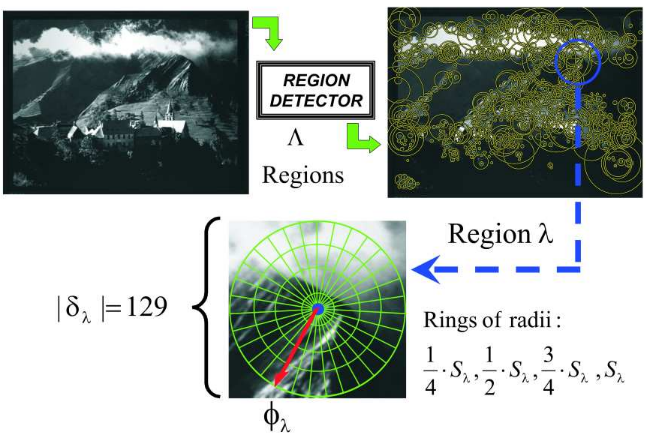

Fig. 3. Descriptor building process. First, a region detector extrats a set Λ of interesting regions. Then, $\forall \lambda \in \Lambda$ we compute a decriptor δ_λ. A descriptor contains the Hölder exponent at the region center (x_λ, y_λ), and of 32 points on the perimeter of four concentric rings, each ring with radii of $\frac{1}{4} \cdot s_\lambda$, $\frac{1}{2} \cdot s_\lambda$, $\frac{3}{4} \cdot s_\lambda$ and s_λ respectively.

from an image. Second, the dominant gradient orientation ϕ_λ is computed, this preserves rotation invariance. Finally, our feature vector δ_λ contains the Hölder exponent α_p of the region center and of 128 concentric points, ordered according to ϕ_λ.

Region extraction: The first step in the process requires stable detection of salient image regions. The type of regions to be extracted will depend on the requirement of the higher level application with respect to invariance. For instance, an interest point detector will suffice when the scale of the imaged scene is not modified. In our work, we use a detector optimized for geometric stability and global point separability, the IPGP2 detector which is the determinant of the Hessian matrix smoothed by a 2D Gaussian kernel [2, 3]. All regions extracted with an interest point detector are assigned the same scale, $w_\lambda = 2.5$ pixels. For images where scale is a factor, we use the Hessian-Laplace detector presented in Ref. [11], which searches for extrema in a linear scale space generated with a Gaussian kernel. After this step we are left with a set Λ of circular regions, the scale is set to $s_\lambda = 5 \cdot w_\lambda$, and w_λ is the scale given by the detector.

Dominant orientation: In order to preserve rotation invariance, the dominant gradient orientation is computed and used as a reference for the subsequent sampling process. For the scale invariant detector, all image regions are normalized to 41x41 bit size using bicubic interpolation. An orientation histogram is constructed using gradient orientations within the interest region, similar to what is described in Ref. [4]. The histogram peak is obtained and thus $\forall \lambda \in \Lambda$ a corresponding dominant orientation ϕ_λ is assigned. In this way, each region is described by a set $\lambda = \{x_\lambda, y_\lambda, s_\lambda, \phi_\lambda\}$, the region center, scale and orientation of the region.

Descriptor: Now that regions are appropriately detected and described with λ, we can now continue to construct our region descriptor $\delta_\lambda, \forall \lambda \in \Lambda$. Our sampling process is simple, see Figure 3, the first element of δ_λ is the Hölder exponent α_p computed at the region center (x_λ, y_λ). Next, the Hölder exponent of points on the perimeter of four concentric rings are sampled, with radii of $\frac{1}{4} \cdot s_\lambda$, $\frac{1}{2} \cdot s_\lambda$, $\frac{3}{4} \cdot s_\lambda$ and s_λ respectively. A total of 32 points on each ring are sampled, starting from the position given by ϕ_λ, uniformly spaced and ordered counterclockwise. Hence, our feature vector δ_λ has 129 dimensions, compared to the 128 of SIFT. The choice of the parameters, such as the size of the rings and the number of sample points, is related with the challenge of building a discriminative descriptor while at the same time maintaining a compact representation. A problem faced by any attempt to describe real-world information. The final values were selected empirically, guided by experimental runs, however an optimization process could be advantageous, i.e. evolutionary computation.

5. Experimental results

In order to effectively evaluate and compare our results, we use standard image sequences provided by the Visual Geometry Group [12]. From each sequence there is one reference image and a set of test images. Due to prior knowledge of the transformation between the reference and test images, we can quantify a matching score. Sample images and experimental results using the following performance metrics are shown in Figure 4. We evaluate with threshold based matching, where two image regions λ_1 and λ_2 are matched if the following relation holds: $d(\delta_{\lambda_1}, \delta_{\lambda_2}) < \delta$. The value of δ is varied to obtain two types of performance curves: one plots *Recall* versus *1-Precision*, characterizing the matching between one test image and the reference image (row 3) [5]; the other, is a double *y-axis* plot, one axis for average *Recall* and the other for average *1-Precision*, that characterizes the performance of the descriptor on a complete sequence (rows 4 & 5). *Recall/1-Precision* gives the number of correct and false matches between two images. *Recall* is the number of correctly matched regions with respect to the number of corresponding regions between two images of the same scene. The number of false matches relative to the total number of matches is represented by *1-Precision*. A perfect descriptor would give a *Recall* equal to 1 for any *Precision*. *Recall* and *1-precision* are defined as in Ref. [5]: $Recall = \frac{\#correct matches}{\#correspondences}$, and $1 - Precision = \frac{\#false matches}{\#correct matches - \#false matches}$. Note that the second type of plot, includes errorbars to visualize the stability of the descriptor. The performance of our descriptor is compared against SIFT. To compute SIFT descriptors, the Harris and Harris-Laplace detectors were used to extract image regions (executables obtained from Ref. [12]). Figure 4 exhibits the following patterns. **Rotation**: the Hölder descriptor outper-

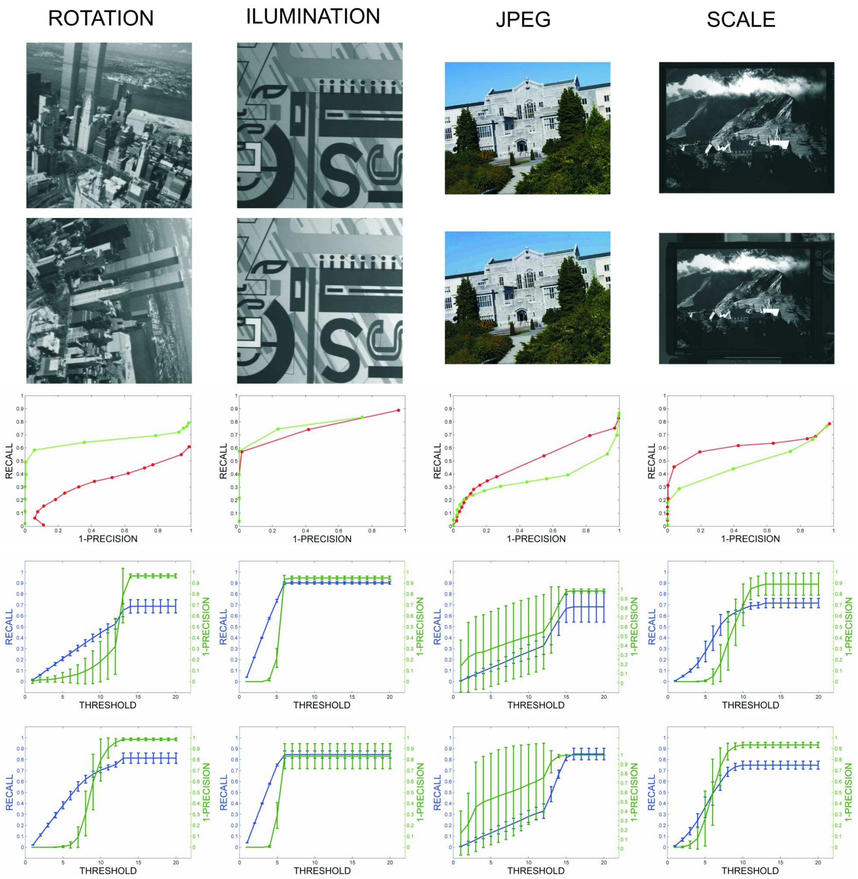

Fig. 4. **Columns**, left to right: 1) Rotation (36 images in sequence), 2) Illumination change (10 images), 3) JPEG compression (6 images), and 4) Scale change (first 6 images of sequence). **Rows**, top to bottom: 1) Reference image, 2) Test Image, 3) Performance between test and reference with Hölder in green and SIFT in red, plotting *Recall* vs. *1-Precision*, 4) & 5) Average performance on the complete image sequence for SIFT and Hölder respectively (*y-axis* with *Recall* in blue and *1-Precision* in green and threshold on *x-axis*).

forms SIFT, with higher *Recall* and better *Precision*. **Illumination & JPEG** : very comparable overall performance. **Scale**: both exhibit the same performance patterns with SIFT consistently better.

6. Conclusions and future work

Results show very promising experimental results, in general we can appreciate how the regularity and SIFT descriptors exhibit comparable performance. For image rotation and illumination change, our Hölder descriptor is consistently better, with the opposite being true for JPEG compression and scale change. In the case of scale change, the performance of our descriptor is expected to be directly related to the method of Hölder exponent estimation. For this reason, an appropriate modification of the oscillations method is necessary. For image compression, it is a consequence of the intrinsic change in image regularity induced by this transformation.