Learning Invariant Region Descriptor Operators with Genetic Programming and the F-Measure

Cynthia B.Perez and Gustavo Olague
CICESE Research Center.
Centro de Investigación Científica y de Educación Superior de Ensenada
Km.107 Carretera Tijuana-Ensenada, 22860. Ensenada, B.C., México.
cbperez@cicese.mx, olague@cicese.mx

Abstract

Recognizing and localizing objects is a classical problem in computer vision that is an important stage for many automated systems. In order to perform object recognition many researchers have focused on local features as the basis of their proposed methodologies. This work is devoted to the task of learning invariant region descriptor operators with genetic programming. The idea is to find a set of expressions that could be equal or better than the weighted gradient magnitude that is normally applied on the SIFT descriptor. This magnitude corresponds to the operator that we would like to improve through genetic programming (GP). The key for a successful problem statement was achieved with the F-measure. After a bibliographical study we have found a criterion that is simple, reliable, and useful in the estimation of such a metric. The measure that we propose here is based on the harmonic mean which is normally used by the information retrieval community. Experimental results show that the evolved descriptor's operator can enhance significantly the overall performance of the SIFT descriptor and surpass other state-of-the-art algorithms.

1 Introduction

The idea of using interest points and region descriptors as a starting stage within a larger vision system has impacted the entire process where identification and matching locations between images is required; i.e., image matching, object recognition and detection, motion tracking and segmentation, 3D reconstruction, image panorama assembly, and epipolar calibration, to name but a few. In particular this paper deals with the synthesis of region descriptor operators through GP.

This work follows our previous research where we have successfully synthesized interest point operators [9, 10, 11]. One of the major conclusions of that work is that a number of interest point operators exist which achieve a high-performance with respect to the repeatability and point dispersion criteria. This is contrary to the common practice of using mainly the Harris operator and its variants as the optimal interest point operator. In this work rather than proposing a novel region descriptor which is definitively a more challenging work, we decide to improve the SIFT algorithm as several researchers have done in the past. For example, Ke and Sukthankar [4] proposed PCA-SIFT in order to represent in a more compact way the normalized gradient patch instead of the standard SIFT representation. Bay et al. [2] designed also a descriptor called SURF, which is a SIFT version specially designed for real time applications. Mikolajczyk and Schmid [6] suggested a descriptor called GLOH, which is a SIFT variant where the local region used to build the histogram is computed with polar coordinates and PCA was also employed. Dalal and Triggs [3] proposed also a SIFT variant called HOG that is based on evaluating well-normalized local histograms of image gradient orientations in a dense grid. Tola et al. [8] introduced a SIFT version named DAISY with the idea of computing fast dense matching in the case of wide baseline configurations using graph-cuts.

In the vein of those previous works we decide to improve also the SIFT algorithm looking for a better description of the information being extracted from the local regions. Our goal is also not to change the whole process but rather change key aspects that could make a greater difference on the global performance. SIFT as described in [5] consists of four major stages: 1) scale space peak detection, 2) keypoint localization, 3) orientation assignment, and 4) keypoint descriptor. We will

focus our discussion on the last stage as our present contribution is devoted to improve the local image descriptor. However, to help the newcomer we recall briefly the main steps. First, potential interest points are detected on the image using the scale space. This is implemented efficiently by constructing a Gaussian pyramid and searching for local peaks (keypoints) in a series of difference-of-Gaussian (DOG) images. Second, candidate keypoints are localized to sub-pixel accuracy and eliminated if found to be unstable. Third, the algorithm identifies the dominant orientations for each keypoint based on its local image patch. The assigned orientation, scale and location for each keypoint enables SIFT to construct a canonical view for the keypoint that is said invariant to similarity transforms. The last stage builds a local image descriptor for each keypoint, based upon the image gradients of a patch of pixels in its local neighborhood. This patch is centered on the keypoint location, rotated with respect to the dominant orientation and scaled to the appropiate size. We propose in this work to name the gradient magnitude normally used in SIFT and its variants as the local image descriptor operator, see Figure 1.
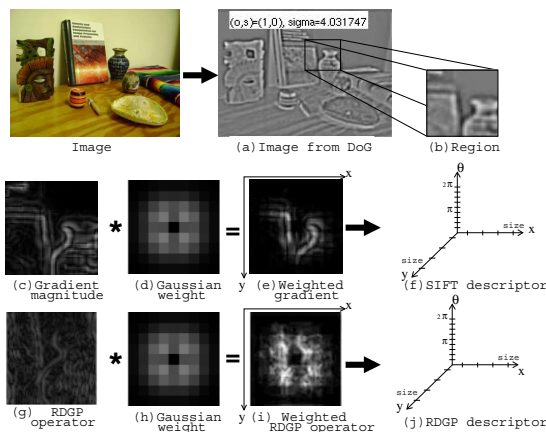
Figure 1. Local image descriptor operator used within SIFT: c) gradient magnitude g) RDGP.

Our idea is to improve such operator through a GP based approach similar to our previous work. However, the challenge here is how to devise a well-posed metric working on the Recall vs 1-precision space.

This paper is organized as follows. Section 2 presents a discussion about our search for a suitable evaluation criterion. Section 3 outlines our GP approach. Section 4 provides implementation details, as well as experimental results and possible future work.

2 Evaluating Regions' Descriptors through the F-measure

Current evaluation techniques described in the local descriptor literature propose metrics that work on the ROC space, as well as on the Recall vs 1-precision space, which are derived from the contingency table. These correspond to graphical evaluations that convey a subjective interpretation specially when plotting curves intersect. Here we propose as an evaluation technique the F-measure in order to obtain a good evaluation methodology to compare local image descriptors. At the time of writing, we are not aware of any work using the F-measure method for evaluating the performance measurement of such tasks. As shown by Agarwal and Roth [1] Recall vs 1-precision space is considered a much better way of reporting and interpreting results because it avoids the computation of the total number of negatives used by ROC curves.

In this way, the overall accuracy could be computed with the following metrics: true negative rate, true positive rate, weighted accuracy, G-mean, precision, recall, and the F-measure. Other measures that could be computed are the area under the ROC or Recall vs 1-precision curves. Another way is simply to address the problem in multiobjective terms and fix one of the coordinates or better to use some evolutionary multiobjective optimization. However, as we will see here the F-measure already provides a measure that is simple, reliable and useful in the estimation of the local descriptor performance. The F-measure is based on the harmonic mean which was originally proposed by the information retrieval community [12]. It gives the best balance between precision and recall metrics. The general formula is defined in the following equation:

$$F_\alpha(p, r) = \frac{(1 + \alpha) \cdot (\mathrm{p} \cdot \mathrm{r})}{(\alpha \cdot \mathrm{p} + \mathrm{r}).} \qquad (1)$$

where p is precision $\{p : 0 \leq p \leq 1\}$, r is recall $\{r : 0 \leq r \leq 1\}$, and $\{\alpha : 0 \leq \alpha \leq \infty\}$. Note that in the case of $\alpha < 1$ the variable with a higher weight is p, while $\alpha > 1$ the variable with a higher weight is r, and when $\alpha = 1$ the precision and recall are well balanced.

3 Evolving Descriptors' Operators with Genetic Programming

Genetic programming is a relatively new and fast developing approach to automatic programming, see Poli et al. [7]. GP is an evolutionary algorithm based methodology inspired from biological evolution to find computer programs that perform a user-defined task. It

is said to be a machine learning technique that automatically solves problems without requiring the user to know or specify the form or structure of the final solution in advance. GP is not typically used to create programs like in the familiar Turing-complete languages that humans normally use for software development. It is instead more common to evolve programs (expressions or formulae) in a more constrained and often domain-specific language. The search space is defined according to our previous analysis, see Trujillo and Olague [10]. The function and terminal sets used in our evolutionary algorithm are the following:

$$F = \left\{ +, |+|, -, |-|, *, \div, \sqrt{I_t}, \right.$$
$$\left. \frac{I_t}{2}, log_2(I_t), D_x G_\sigma, D_y G_\sigma, G_\sigma \right\} \quad (2)$$
$$T = \{I, D_x, D_{xx}, D_{yy}, D_{xy}, D_y\}$$

where I is the input image and I_t can be any of the terminals in T, as well as the output of any of the functions in F; D_u symbolizes the image derivatives along direction u then $D_u = I * G_{u(\sigma=1)}$; G_σ are the Gaussian smoothing filters with $\sigma = 1$ or 2; $D_u G_\sigma$ represents the derivative of a Gaussian filter with image blur σ. An appropriate fitness function is decisive to the GP process success. Thus, our fitness function is based on a well balance precision and recall data as explained earlier:

$$Q = \text{argmax} \left\{ F_\alpha(P_i^x, R_i^x) = \sum_{i=1}^n \frac{(1+\alpha) \cdot (p_i \cdot r_i)}{(\alpha \cdot p_i) + r_i} \right\}$$
$$where \quad Q : F_\alpha(P^s, R^s) \quad \geq \quad F_\alpha(P^t, R^t)$$
$$(3)$$

with n representing the number of thresholds. Precision data from an image pair are denoted by $P^x = (p_1, p_2, ..., p_n)$ and recall data by $R^x = (r_1, r_2, ..., r_n)$. The ranking order of Q is ascendent where the highest value of Q corresponds to the descriptor with the best performance. It is also possible to use the mean of Q instead of the total sum of F_α.

4 Experimental Results

In this section, we introduce experimental results for the local descriptors' evaluation. The proposed approach was implemented as in general machine learning through two basic stages: training and testing which were programmed on Matlab with the GP toolbox (GPLAB[1]). The core platform of our descriptor algorithm is based on SIFT features, programmed in Matlab/C[2]. For training we have used the boat and Leuven

---

[1] http://gplab.sourceforge.net/index.html
[2] http://vision.ucla.edu/ vedaldi/code/sift/sift.html

image pairs, which have rotation, scale and illumination transformations. For testing, five image pairs were used with different image transformations. All image sets were downloaded from the Visual Geometry Group website[3], along with Matlab source code to compute precision and recall data. For more details of the dataset used in this paper, see [6]. The training test could seem to be rather small to learn something useful and general. Note that we are learning a mathematical expression based on geometric and photometric principals without any kind of semantic interpretation. According to our previous work it is only important to train with an example of the major transformation being studied. Its generality could be tested with other image sequences keeping the main hypothesis as is normally done; for example, with the Harris operator. Results are shown in Figure 3 where it is easily to observe the superiority of the RDGP operator compared with the GLOH and SIFT descriptors. In those graphs we decide to change $\alpha = 0.5$ except for Figure 3(c) where was used $\alpha = 2$. Note that our space is now r vs $1 - p$ and the domain is defined as follows: when $\alpha < 1$ the variable with the higher weight is $1 - p$ then is logical to observe that we are separating the curves along this axis. On the other hand, when $\alpha > 1$ then we are separating the curves along the recall axis. The discovered RDGP, see Fig. 2, could be seen as a geometric mean, which indicates the central tendency or typical value of a set of numbers and it only applies to positive numbers. In a special case the geometric mean is also the arithmetic-harmonic mean.
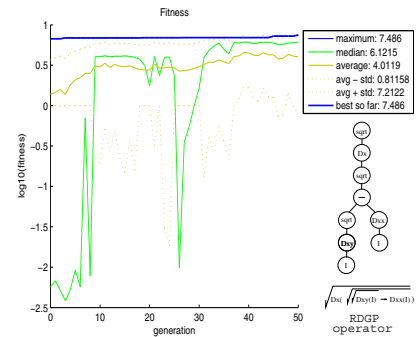
Figure 2. Statistics for RDGP showing fitness evolution and best individual.

5 Conclusions

In this paper, we have proposed a new methodology based on GP with the goal of synthesizing region

---

[3] http://www.robots.ox.ac.uk/vgg/research/

Table 1. RDGP algorithm set up.

Parameters	Description
Generations	50
Population size	50 individuals
Initialization	Ramped Half-and-Half
Crossover	0.90
Mutation	0.10
Selection	Stochastic Universal Sampling
Survival	Keep Best Survival Strategy
Fitness	$\alpha = 1$

descriptor operators that actually achieves excellent results. The quantitative measurement presented in this paper was the key to pose properly the evolutionary process; since, the measurement is basic to make a correct judgment on the performance achieved by each solution. In future work we would like to explore our methodology in real-world recognition problems.

References

[1] S. Agarwal and D. Roth. Learning a sparse representation for object detection. *ECCV*, 2353:97–101, 2002.

[2] H. Bay, T. Tuytelaars, and L. V. Gool. Surf: Speeded up robust features. *ECCV*, 3951:404–417, 2006.

[3] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR - Volume 1*, pages 886–893, Washington, DC, USA, 2005.

[4] Y. Ke and R. Sukthankar. Pca-sift: A more distinctive representation for local image descriptors. In *CVPR, 27 June - 2 July, Washington, DC*, volume 2, pages 506–513. IEEE Computer Society, 2004.

[5] D. Lowe. Object recognition from local scale-invariant features. *ICCV*, pages 1150–1157, 1999.

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

[7] R. Poli, W. B. Langdon, and N. F. McPhee. *A field guide to genetic programming*. Published via http://lulu.com and freely available at http://www.gp-field-guide.org.uk, 2008. (With contributions by J. R. Koza).

[8] E. Tola, V. Lepetit, and P. Fua. A fast local descriptor for dense matching. In *CVPR, 23-28 June, Anchorage, AK*. IEEE Computer Society, 2008.

[9] L. Trujillo and G. Olague. Synthesis of interest point detectors through genetic programming. In *GECCO, Seattle, WA*, volume 1, pages 887–894. ACM, 2006.

[10] L. Trujillo and G. Olague. Using evolution to learn how to perform interest point detection. In *ICPR, Hong Kong, China*, volume 1, pages 211–214, 2006.

[11] L. Trujillo and G. Olague. Automated design of image operators that detect interest points. *Evolutionary Computation*, 2008. to appear.

[12] C. Van-Rijsbergen. *Information Retrieval*. Butterworth-Heinemann, 1979.
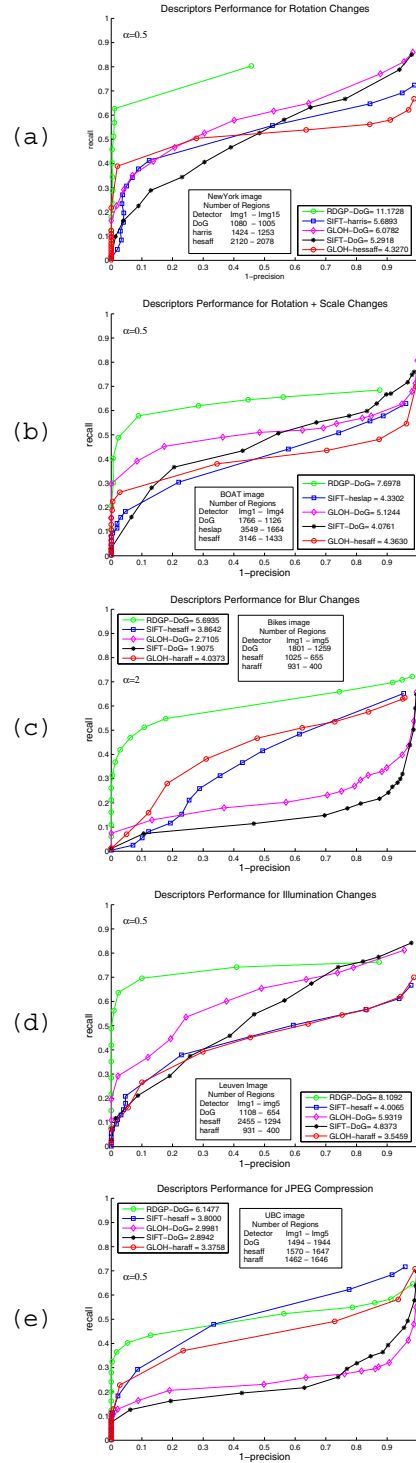
Figure 3. Experimental results for descriptors performance under different image transformations.