6th ANNUAL "HUMIES" AWARDS

Evolutionary Learning of Local Descriptor Operators for Object Recognition

Present :

Cynthia B. Pérez

and

Gustavo Olague

EvoVisión Laboratory
Computer Science Department,
CICESE Research,
Ensenada B.C. México

July 2009

Related Publications

- Perez C.B., Olague G.

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

  International Conference on Pattern Recognition (ICPR).

  December 8-11, 2008.

- Perez C.B., Olague G.

  "Evolutionary Learning of Local Descriptor Operators for Object Recognition".

  Genetic and Evolutionary Computation Conference (GECCO).

  July 8-12, 2009.

Human Competitiveness

This work fullfils 7 of the 8 criteria for human competitiveness:

(A) The result was patented as an invention in the past, is an improvement over a *patented invention*, or would qualify today as a patentable new invention.

(B) The result is equal to or better than a result that was accepted as a new scientific result at the time when it was published in a peer-reviewed scientific journal.

(C) The result is equal to or better than a result that was placed into a database or archive of results maintained by an internationally recognized panel of scientific experts.

(D) The result is publishable in its own right as a new scientific result 3/4 independent of the fact that the result was mechanically created.

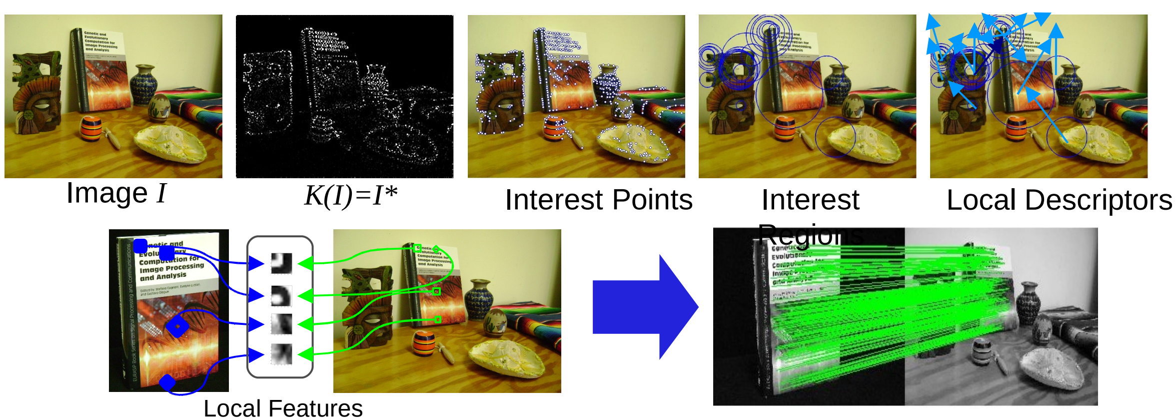
(E) The result is equal to or better than the most recent human-created solution to a long-standing problem for which there has been a succession of increasingly better human-created solutions.

(F) The result is equal to or better than a result that was considered an achievement in its field at the time it was first discovered.

(G) The result solves a problem of indisputable difficulty in its field.

The Problem

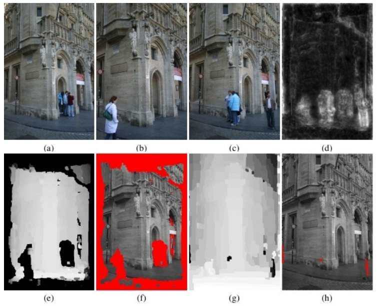
- The computer vision (CV) problem addressed in this work is, **Invariant Local Descriptors**.
- Local descriptors extracted from interest regions have impacted to the CV community due to its simplified methodology for CV applications.
- The idea of using local features in the context of matching and recognition under different viewing conditions was first proposed by Schmid and Mohr[1].



[1]C.Schmid and R.Mohr. Local grayvalue invariants for image retrieval. IEEE PAMI. 19(5): 530-534. 1997.

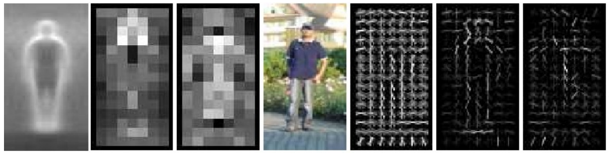
Wide Range of Applications

- Object Recognition [1-6]
- Image Retrieval [7-10]
- Human Detection [11]
- Texture Classification [9,12,13]
- 3D Reconstruction [14,15]
- Motion Field Prediction [16]
- Image Deformation [17,18]
- Image Panoramic Assembly [19]
- Face Detection [13,20]

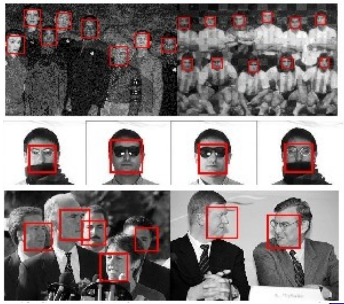


3D RECONSTRUCTION

MOTION FIELD PREDICTION



HUMAN DETECTION

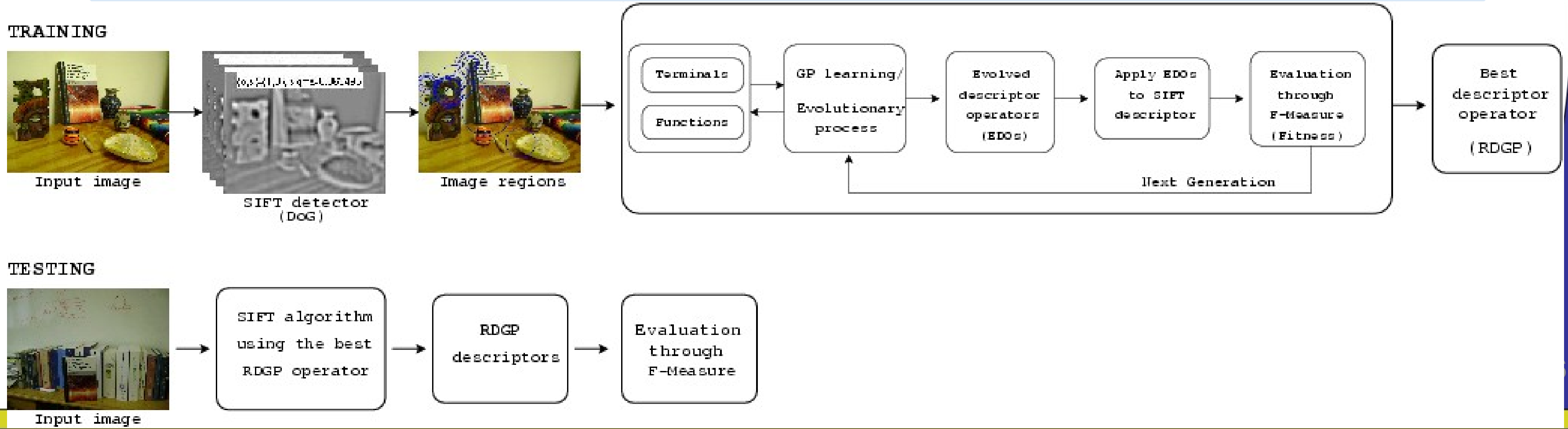


FACE DETECTION

IMAGE PANORAMIC ASSEMBLY

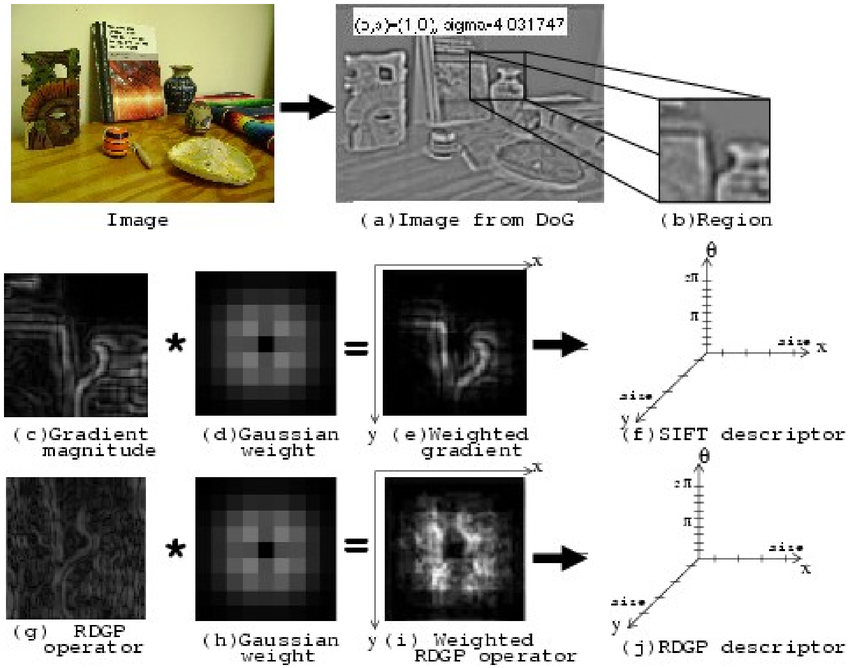
Our Approach

- Invariant local descriptor is posed as an optimization problem.
- GP is used to synthesized mathematical expressions that are used to improve the patented SIFT descriptor.
- The results are called RDGPs (Region Descriptor with Genetic Programming).
- The F-Measure is proposed as a adequate fitness function as well as a measure for the performance evaluation of local descriptors.
- A widely accepted testbed is used in the evaluation.
- The proposed descriptor is tested in an object recognition application.



Motivation for designing RDGPs

- Development a technique that is simple, automated and reliable for improving local descriptors.
- Better descriptor performance, better real applications.



Performance Evaluation: F-Measure

- This measure gives the best balance between precision and recall metrics commonly used in graphs to evaluate local descriptors[2].
- We claim that the F-Measure gives a better interpretation of the results than only plotting them.

TestBed:

- INRIA Rhone Alpes
- University of Oxford
- Katholieke Universiteit Leuven
- Center of Machine Perception at the Czech Technical University

[2]K.Mikolajczyk and C.Shmid. A performance Evaluation of Local Descriptors IEEE PAMI. 27(10):1615-1630. 2005.

Performance Evaluation: F-Measure

ROTATION

NewYork

ILLUMINATION

Leuven

ROTATION + SCALE

Boat

ROTATION + SCALE

Bark

BLUR

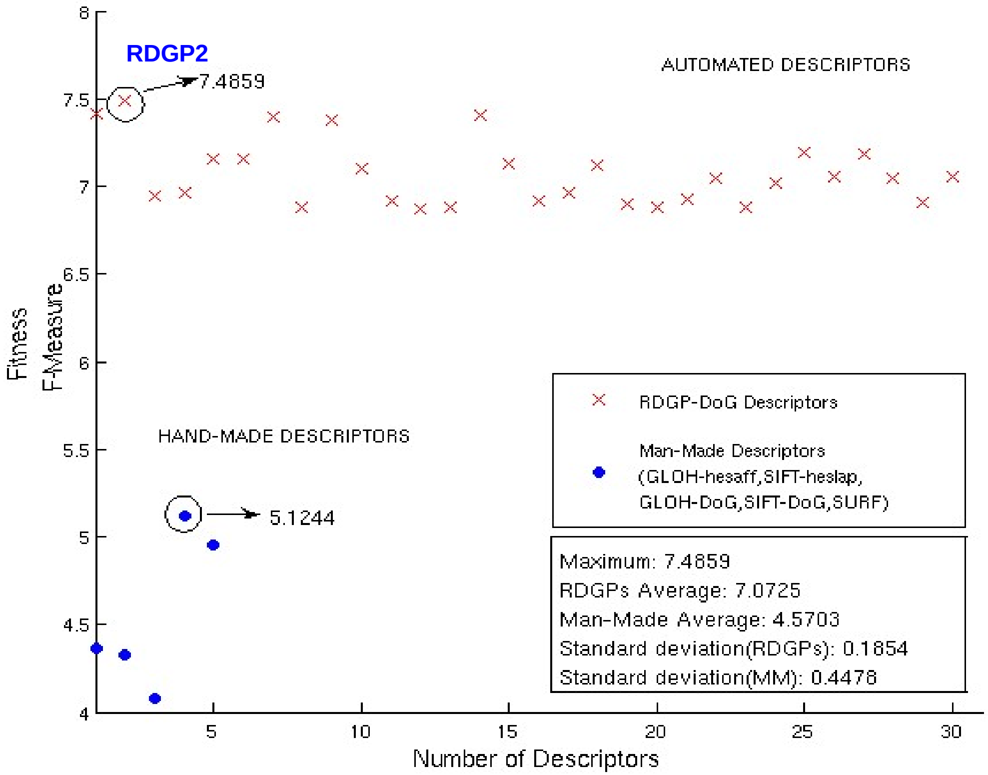
Trees

JPEG COMPRESSION

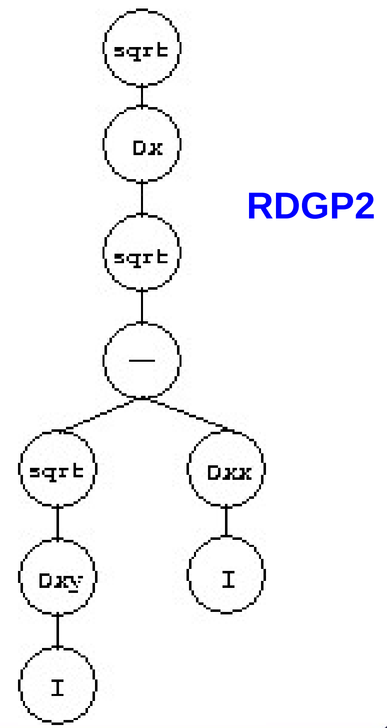
UBC

Results

- Our approach produced 30 RDGPs that outperformed all the state-of-art descriptors published with the same testbed.



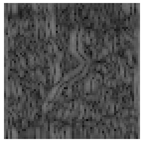
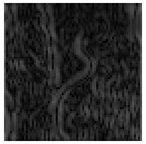
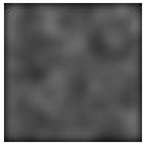
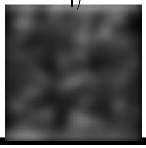
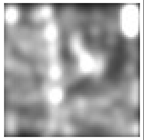
Best Result



Results

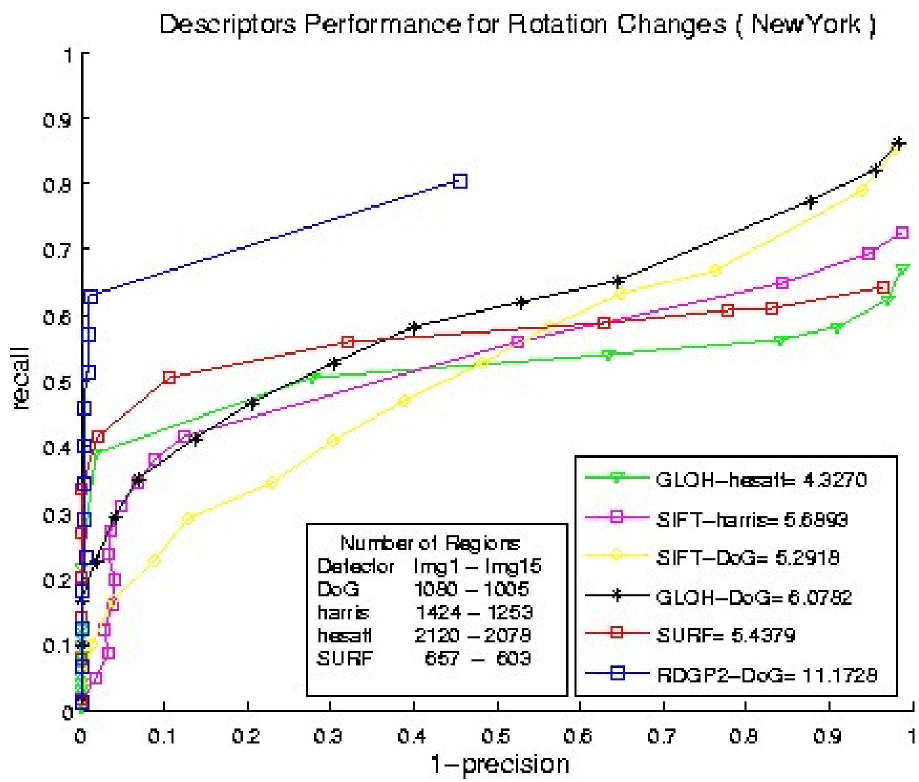
The 5 Best Evolved RDGPs

Original image region

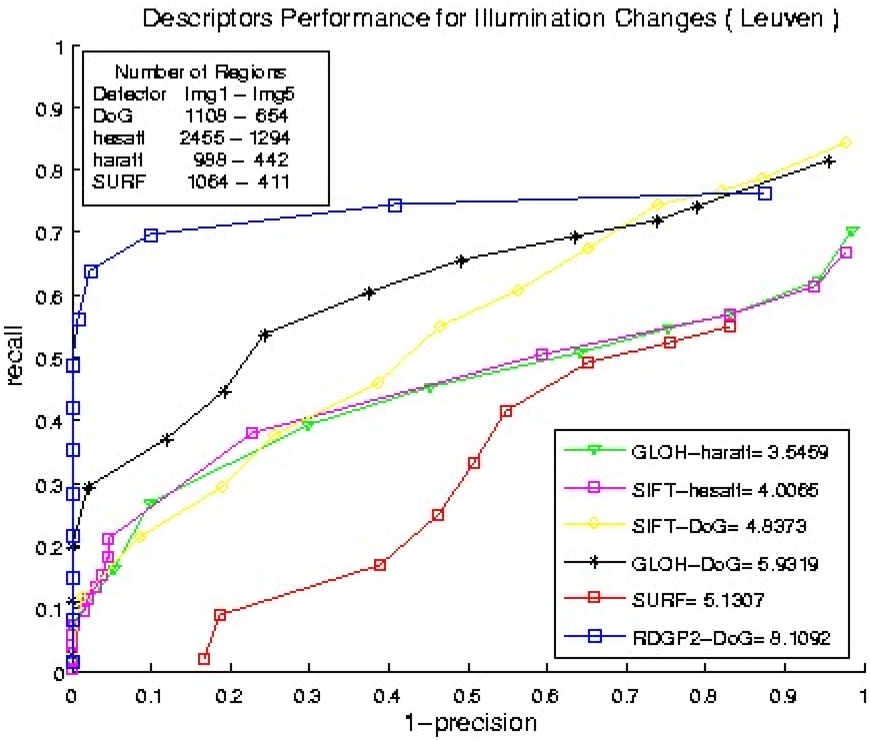
Descriptor	Fitness	Individual's Expression	Mathematical Expression	Image Region
RDGP1	7.4158	sqrt(sqrt(Dx(sqrt(Dxx(Image)))))	$\sqrt{\sqrt{D_x\sqrt{D_x\sqrt{D_{xy}(I)}}}}$	
RDGP2	7.4859	sqrt(Dx(sqrt(substract(sqrt(Dxy(image)),Dxx(image)))))	$\sqrt{D_x\sqrt{\sqrt{D_{xy}(I)}-D_{xx}(I)}}$	
RDGP3	7.1812	Gauss2(Gauss2(sqrt(Dx(Dy(Dx(Dx(image)))))))	$G_{\sigma=2}(G_{\sigma=2}(\sqrt{D_{xy}(D_{xx}(I))}))$	
RDGP4	7.3928	Gauss2(absdif(Gauss2(absdif(absdif(Dx(image),Dx(Dx(image))),Dx(Logarithm(Dxx(image)))))),Half(Dx(Dy(image)))))	$G_{\sigma=2}\left\lvert G_{\sigma=2}\left(\left\lvert\left\lvert D_y(I)-D_{xx}(I)\right\rvert-D_y(\log(D_{xx}(I)))\right\rvert\right)-\frac{D_{xx}(I)}{2}\right\rvert$	
RDGP5	7.4053	Gauss1(sqrt(Gauss2(sqrt(sqrt(substract(sqrt(Gauss1(Dx(image))),divide(Dxx(image),absadd(Dx(image),Dy(image)))))))))	$G_{\sigma=1}\sqrt{G_{\sigma=2}\sqrt{\sqrt{G_{\sigma=1}D_y(I)}-\frac{D_{xx}(I)}{\lvert D_x(I)+D_y(I)\rvert}}}$	

Results

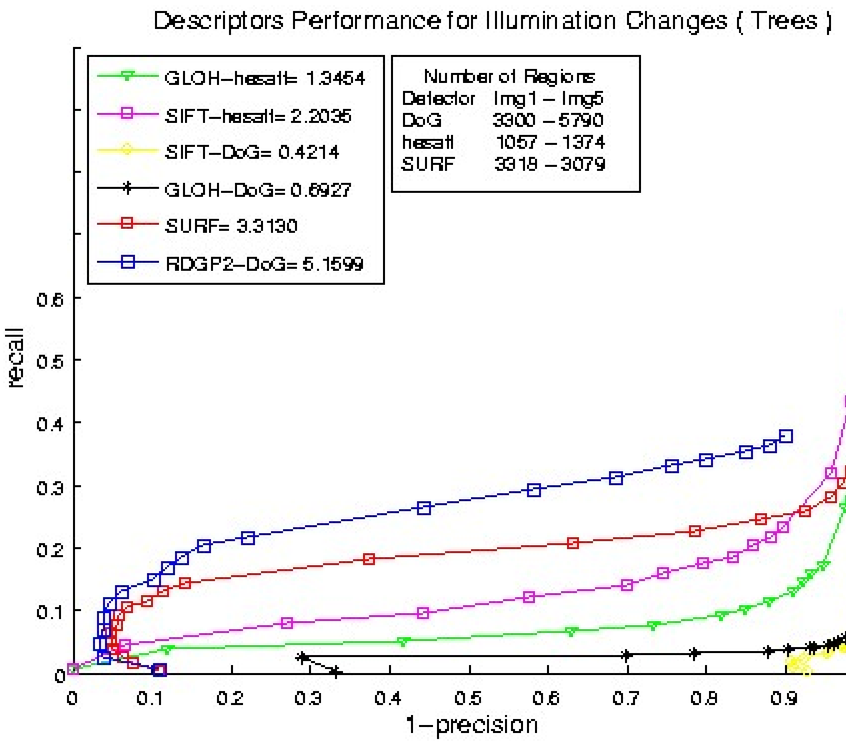
- We obtained much better performance that the human-made descriptor algorithms.



Results

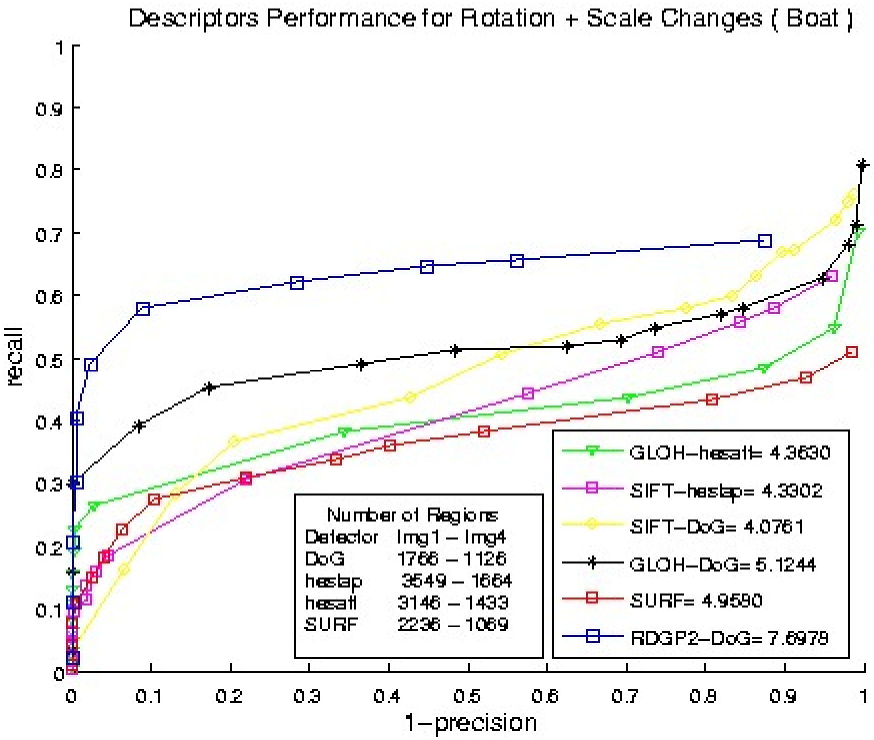


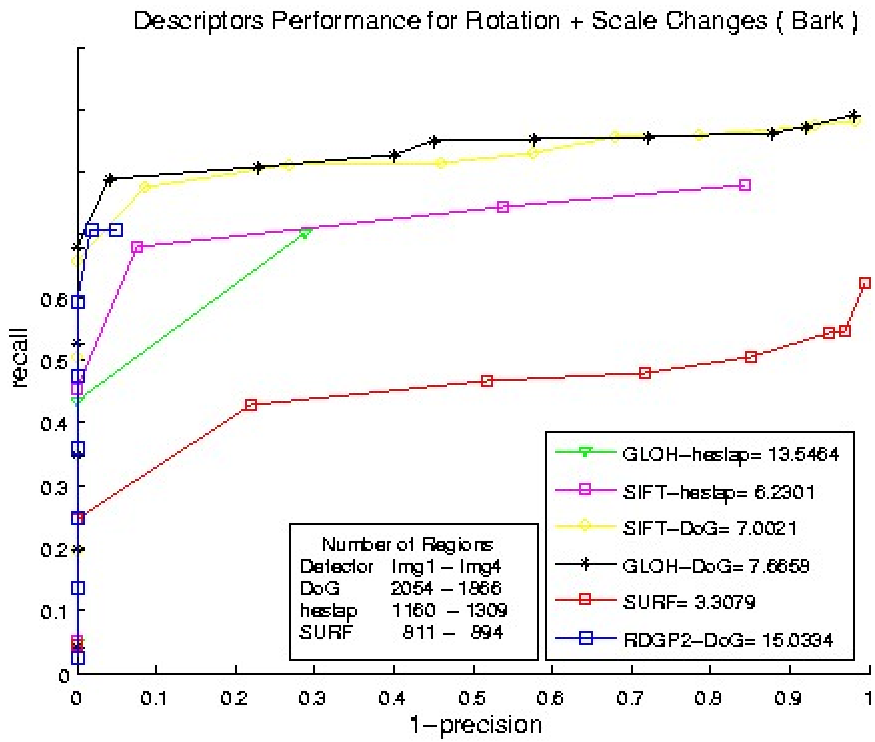
Descriptors Performance for Illumination Changes (Leuven)
Number of Regions
Detector Img1 – Img5
DoG 1109 – 654
hesaff 2455 – 1294
haraff 988 – 442
SURF 1064 – 411
GLOH–haraff= 3.5459
SIFT–hesaff= 4.0065
SIFT–DoG= 4.8373
GLOH–DoG= 5.9319
SURF= 5.1307
RDGP2–DoG= 8.1092
recall
1–precision



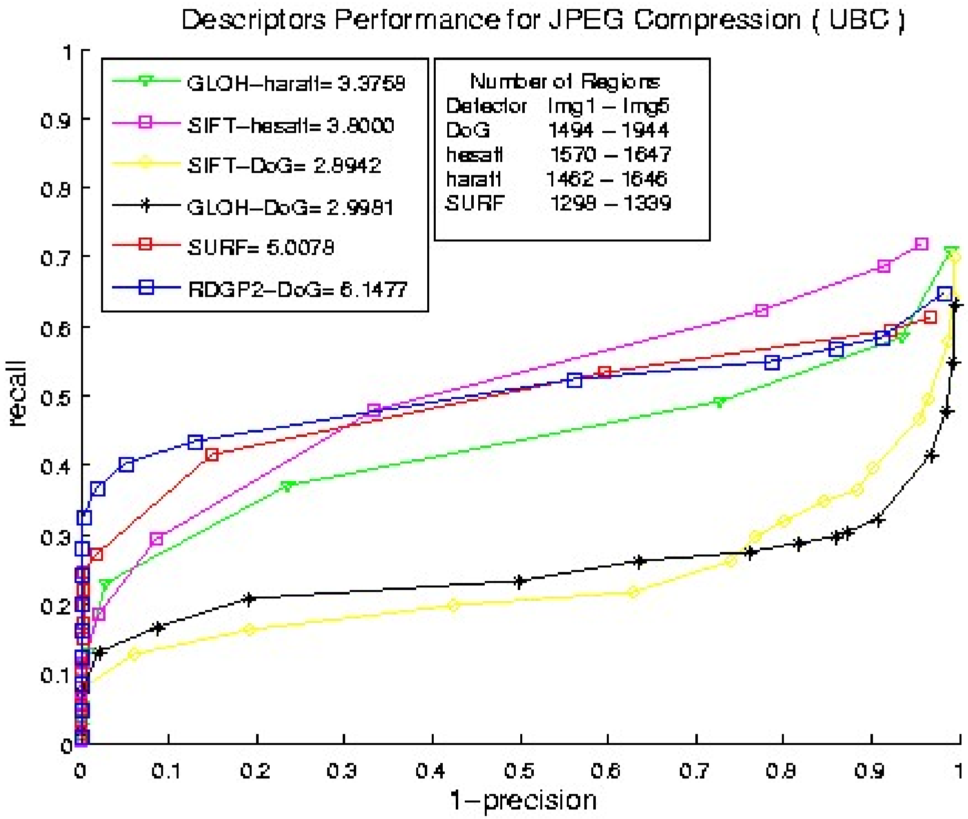
Descriptors Performance for Illumination Changes (Trees)
GLOH–hesaff= 1.3454
SIFT–hesaff= 2.2035
SIFT–DoG= 0.4214
GLOH–DoG= 0.6927
SURF= 3.3130
RDGP2–DoG= 5.1599
Number of Regions
Detector Img1 – Img5
DoG 3300 – 5790
hesaff 1057 – 1374
SURF 3318 – 3079
recall
1–precision

Results


Descriptors Performance for Rotation + Scale Changes (Boat)
recall
1−precision
Number of Regions
Detector Img1 – Img4
DoG 1766 – 1126
heslap 3549 – 1664
hesaff 3146 – 1433
SURF 2236 – 1069
GLOH−hesaff= 4.3630
SIFT−heslap= 4.3302
SIFT−DoG= 4.0761
GLOH−DoG= 5.1244
SURF= 4.9580
RDGP2−DoG= 7.6978


Descriptors Performance for Rotation + Scale Changes (Bark)
recall
1−precision
Number of Regions
Detector Img1 – Img4
DoG 2054 – 1866
heslap 1160 – 1309
SURF 811 – 894
GLOH−heslap= 13.5464
SIFT−heslap= 6.2301
SIFT−DoG= 7.0021
GLOH−DoG= 7.6659
SURF= 3.3079
RDGP2−DoG= 15.0334

Results

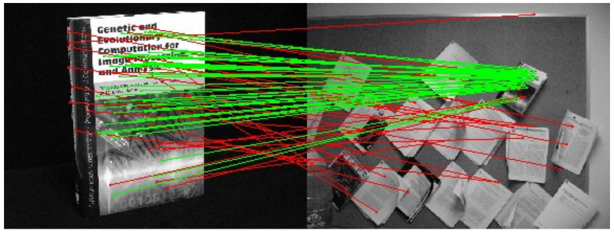


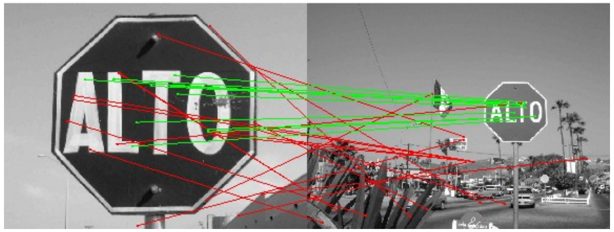
Descriptors Performance for JPEG Compression (UBC)
GLOH−harafl= 3.3758
SIFT−hesafl= 3.8000
SIFT−DoG= 2.8942
GLOH−DoG= 2.9981
SURF= 5.0078
RDGP2−DoG= 6.1477
Number of Regions
Detector Img1 − Img5
DoG 1494 − 1944
hesafl 1570 − 1647
harafl 1462 − 1646
SURF 1298 − 1339
recall
1−precision
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Object Recognition

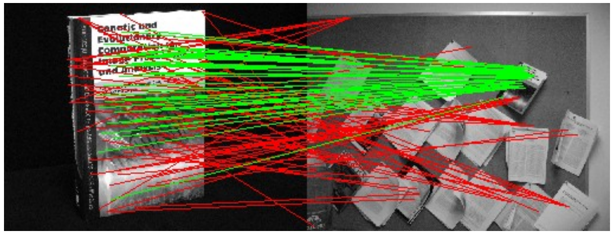
INDOOR SCENARIOS

OUTDOOR SCENARIOS

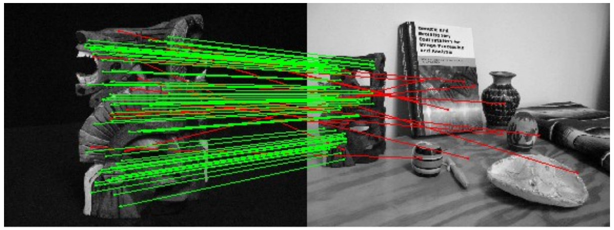


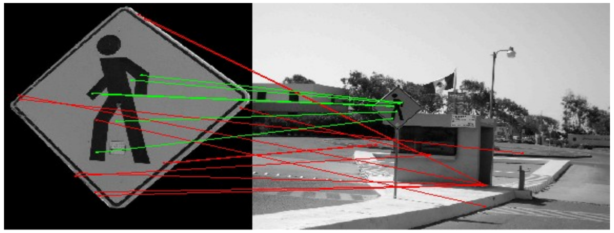


← RDGP2 →

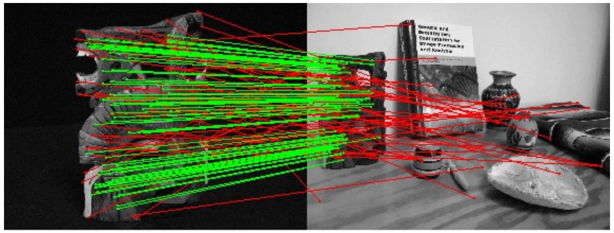


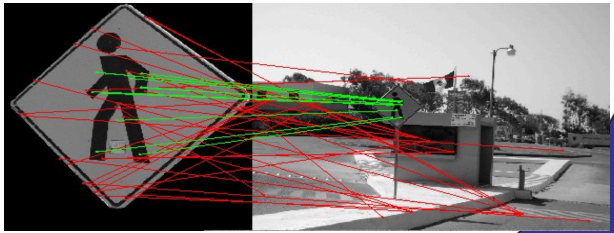
← SIFT →





← RDGP2 →





← SIFT →

Human Competitiveness

Why should this work win?

- The results obtained in this work fulfills 7 of the 8 human competitive criteria.
- Our methodology for automatically obtaining new descriptor operators using GP represents a new approach within the CV community.
- We believe that this kind of formulation shows a rigorous path in the design of computer vision applications where GP plays a major role; thus, strengthening the emerging area of evolutionary computer vision.

Human Competitiveness

(A) The result was patented as an invention in the past, is an improvement over a *patented invention*, or would qualify today as a patentable new invention.

Our proposed methodology for synthesizing descriptor operators represent an improvement over a patented descriptor algorithm called SIFT (Scale Invariant Feature Transform).

The **SIFT patent** is the following:
"***Method and apparatus for identifying scale invariant features in an image and use of same for locating an object in an image"***. David G. Lowe, US Patent 6,711,293 (March 23, 2004). Asignee: The University of British Columbia.

Human Competitiveness

(B) The result is equal to or better than a result that was accepted as a new scientific result at the time when it was published in a peer-reviewed scientific journal.

Here, we compared our results with previous published descriptors from which their evaluation technique was based on a recall vs 1-precision space. Thus, we tested several works to compare our descriptor algorithm and in particular we found that our results surpassed the overall performance of previous local descriptors including the following:

- David G. Lowe, ***"Distinctive image features from scale-invariant keypoints,"*** International Journal of Computer Vision, 60(2):91-110, 2004.
- K. Mikolajczyk, C. Schmid, ***A performance evaluation of local descriptors***. IEEE Transactions on Pattern Analysis and Machine Learning, 27(10):1615-1630, 2005.
- Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool, "***SURF: Speeded Up Robust Features"***, Computer Vision and Image Understanding (CVIU), 110(3):346-359, 2008.

Human Competitiveness

(C) The result is equal to or better than a result that was placed into a database or archive of results maintained by an internationally recognized panel of scientific experts.

We used a testbed that is widely accepted as a standard performance evaluation for local descriptors in the computer vision community. It is available at the following address:

http://www.robots.ox.ac.uk/~vgg/research/affine/

Human Competitiveness

(D),(E) and (F)

(D) The result is publishable in its own right as a new scientific result 3/4 independent of the fact that the result was mechanically created.

(E) The result is equal to or better than the most recent human-created solution to a long-standing problem for which there has been a succession of increasingly better human-created solutions.

(F) The result is equal to or better than a result that was considered an achievement in its field at the time it was first discovered.

Our methodology for automatically obtaining new descriptor operators using genetic programming represents a new approach within the computer vision field; in particular, it address a new approach where local descriptors could be synthesized through GP. As a by product, the results found by genetic programming in the experimental stage surpassed our initial expectations; indeed, we obtained much better performance than the human-made descriptor algorithms. As a conclusion, we have improved the SIFT algorithm which has been considered until now, an achievement in its field using GP.

(G) The result solves a problem of indisputable difficulty in its field.

Today, most computer vision conferences and journals devote a special session or section to local descriptors research because it has became a powerful technique for solving real-world vision problems. Thus, *our proposed technique opens a research avenue towards evolutionary learning of local descriptors*. Here, we demostrated the effectiveness of our GP approach through an extensive experimental study and its application using an object recognition problem.

References

[1] D.G. Lowe. Object recognition from local scale-invariant features. In Proceedings of the IEEE Conference on Computer Vision. pp. 1150 -- 1157. 1999.

[2] S. Lazebnik, C. Schmid and J. Ponce. Semi-local affine parts for object recognition. In Proceedings of the BMVC. Vol. 2, pp. 959 -- 968. 2004.

[3] A. Stein and M. Hebert. Incorporating background invariance into feature-based object recognition. IEEE Workshops on Application of Computer Vision. pp. 37 -- 44. 2005.

[4] E. Mortensen, H. Deng and L. Shapiro. A SIFT descriptor with global context. In Proceedings of the IEEE Conference on CVPR. Vol. 1, pp. 184 -- 190. 2005.

[5] H. Bay, B. Fasel and L. Van Gool. Interactive museum guide: fast and robust recognition of museum objects. In Proceedings of the first International Workshop on Mobile Vision. 2006.

[6] J. Geusebroek. Compact object descriptors from local colour invariant histograms. In British Machine Vision Conference. Vol. 3, pp. 1029 -- 1038. 2006.

[7] Yan Ke, Rahul Sukthankar. “PCA-SIFT: A More distinctive representation for local image descriptors”. Computer Vision and Pattern Recognition, CVPR 2004

References

[8] S. Lazebnik, C. Schmid and J. Ponce. A Sparse Texture Representation Using Affine-Invariant Regions. CVPR. Vol. 2, pp. 319 -- 324. 2003.

[9] B.S. Manjunath, J. Ohm, V. Vasudevan and A. Yamada. Color and texture descriptors. IEEE Transactions on Circuits and Systems for Video Techonology. 11(6): 703 -- 715. 2001.

[10] A. Carkacioglu and F. Yarman-Vural. SASI: a generic texture descriptor for image retrieval. Pattern Recognition. 33(11): 2615 -- 2633. 2003.

[11] Navneet Dalal and Bill Trigs. Histograms of oriented gradients for human detection. In Proceedings of the IEEE CVPR. pp. 886 -- 893. 2006.

[12] A. Bosch, A. Zisserman, X. Munoz. Representing shape with a spatial pyramid kernel. International Conference on Image and Video Retrieval. pp. 401 -- 408. 2007.

[13] J. Chen, S. Shan, G. Zhao, X. Chen, W. Gao and M. Pietikainen. A robust descriptor based on Weber's law. CVPR 2008.

[14] H. Bay, T. Tuytelaars and L.Van Gool. SURF: Speeded up robust features. In Proceedings of the European Conference on Computer Vision. LNCS 3951, pp.404 -- 417. 2006.

References

[15] E. Tola, V. Lepetit and P. Fua. A fast descriptor for dense matching. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2008.

[16] C. Liu, J. Yuen, A. Torralba and J. Sivic. SIFT flow: dense correspondence across different scenes. European Conference on Computer Vision. Marseille, France. October 2008.

[17] H. Ling and D. Jacobs. Deformation invariant image matching. In Proceedings on the International Conference on Computer Vision. Vol. 2. pp. 1466 -- 1473. 2005.

[18] H. Cheng, Z. Liu, N. Zheng and J. Yang. A deformable local image descriptor. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2008.

[19] M. Brown. R.Szeliski and S. Winder. Multi-image matching using multi-scale oriented patches. Conference on Computer Vision and Pattern Recognition. pp. 510 -- 517. 2005.

[20] S. Sarfraz and O. Hellwich. Head pose estimation in face recognition across pose scenarios. International Conference on Computer Vision Theory and Applications. pp. 235 -- 242. 2008.

Table 1
Overview of local Descriptors.

Descriptor	*Operator*	*Organization*	*Application*	*Inspired on SIFT*
SIFT [3]	$\| \nabla\phi \|= \sqrt{\phi_x^2 + \phi_y^2}$	3D histogram	Object detection	✓
PCA-SIFT [4]	$\| \nabla\phi \|= \sqrt{\phi_x^2 + \phi_y^2}$	Directly	Image retrieval	✓
RIFT [5]	$\| \nabla\phi \|= \sqrt{\phi_x^2 + \phi_y^2}$	2D histogram	Texture image retrieval and classification [5] Butterflies recognition [6]	✓
GLOH [9]	$\| \nabla\phi \|= \sqrt{\phi_x^2 + \phi_y^2}$	3D histogram	-	✓
BSIFT [10]	$I^{(k+1)}(x,y) = I^{(k)}(x,y) + \tau\nabla^2 I^{(k)}(x,y)$	3D histogram	Object detection	✓
SIFT-GC [11]	$\| \nabla\phi \|= \sqrt{\phi_x^2 + \phi_y^2}$	3D histogram	-	✓
CSIFT [12]	$\hat{E}_{\lambda' w} = \sqrt{\hat{E}^2_{\lambda' x} + \hat{E}^2_{\lambda' y}}$	3D histogram	-	✓
SURF [13]	Haar Wavelet * G_σ	Directly	3D reconstruction [13] Museum objects recog. [25]	✓
HOG [14]	$\| \nabla\phi \|= \sqrt{\phi_x^2 + \phi_y^2}$	Cell-blocks	Human detection	✓
PHOG [15]	$\| \nabla\phi \|= \sqrt{\phi_x^2 + \phi_y^2}$	2-D histogram	Image classification	✓
DAISY [16]	G_Σ from $G_o^\Sigma = G_\Sigma * \left(\frac{\partial I}{\partial o}\right)^+$	Orientation maps	3D reconstruction	✓
SIFT-flow [18]	$\| \nabla\phi \|= \sqrt{\phi_x^2 + \phi_y^2}$	3D histogram	Motion field prediction	✓
Steerable filters [26]	$E_n(\theta) = [G_n^\theta]^2 + [H_n^\theta]^2$	Directly	Texture analysis	×
HTD-TBD-EHD [30] SCD-CSD-CLD-DCD	Bank of filters HSV,HMMD & YCrCb	Histogram	Texture classification Image retrieval	×
SASI [27]	-	Clique windows	Image retrieval	×
GIH [32]	I(x,y)	Geodesic-Intensity Histogram	Synthetic deformation Non-affine deformation	×
MOPs [17]	Haar Wavelet Transform	directly	Panoramic image stitching	×
Wiccest [29]	$\hat{E}_{\lambda\lambda m}(x,y,\sigma_i) = G_{\lambda\lambda}(x,y,\sigma_i) * E_m(x,y)$	Histogram	Object recognition	×
Hölder [22]	$\alpha_p = sup_s\{f \in C^s(x_o)\}$	Concentric rings	Image Matching	×
LESH [31]	Local Energy Model	2D histogram	Face recognition	×
SMD [33]	I(x,y)	Stable pairs	Image matching	×
DLID [23]	$\| \nabla\phi \|= \sqrt{\phi_x^2 + \phi_y^2}$	2D histograms	Image deformations	×
WLD [24]	$\xi(I_c) = arctan\left[\sum_{i=0}^{p-1}\left[\frac{I_i - I_c}{I_c}\right]\right]$	2D histograms	Texture classification Face detection	×