HAL
open science

Multiscale Shape Description with Laplacian Profile and Fourier Transform

Evanthia Mavridou, James L. Crowley, Augustin Lux

To cite this version:

Evanthia Mavridou, James L. Crowley, Augustin Lux. Multiscale Shape Description with Laplacian Profile and Fourier Transform. International Conference on Image Analysis and Recognition, ICIAR 2014, Oct 2014, Vilamoura, Algarve, Portugal. pp.46 - 54, 10.1007/978-3-319-11758-4_6 . hal-01079658

HAL Id: hal-01079658

https://inria.hal.science/hal-01079658v1

Submitted on 3 Nov 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Multiscale Shape Description with Laplacian Profile and Fourier Transform

Evanthia Mavridou[1,2], James L. Crowley[1,2], and Augustin Lux[1,2]

[1]Univ. Grenoble Alpes, LIG, F-38000 Grenoble, France
[2]Inria Grenoble Rhône-Alpes Research Centre, LIG, F-38000 Grenoble, France
{evanthia.mavridou,james.crowley,augustin.lux}@inria.fr

Abstract. We propose a new local multiscale image descriptor of variable size. The descriptor combines Laplacian of Gaussian values at different scales with a Radial Fourier Transform. This descriptor provides a compact description of the appearance of a local neighborhood in a manner that is robust to changes in scale and orientation. We evaluate this descriptor by measuring repeatability and recall against 1-precision with the Affine Covariant Features benchmark dataset and as well as with a set of textureless images from the MIRFLICKR Retrieval Evaluation dataset. Experiments reveal performance competitive to the state of the art, while providing a more compact representation.

Keywords: robust image description, scale invariance, local appearance description, compact descriptor, variable vector length

1 Introduction

Robustness is fundamental for image description. While discriminative power has always been an important consideration, applications have increasingly imposed constraints on robustness, memory requirements and computation cost. In an effort to provide a compact robust descriptor, we have explored a new visual descriptor based on combining a Laplacian Profile with a Radial Discrete Fourier Transform (LP-RDFT) [13]. An interesting property of the LP-RDFT is that it has a adjustable description length, making it possible to tradeoff description length for discriminative power. In comparison with popular descriptors, we have found that at equivalent discrimination levels, the LP-RDFT descriptor provides a much smaller description length. We have also found that the LP-RDFT provides useful discrimination at even its smallest vector lengths [13].

To evaluate the utility of LP-RDFT, we compared its performance for repeatability and recall against 1-precision to other descriptors chosen from the state of the art using the Affine Covariant Features benchmark dataset and the MIRFLICKR Retrieval Evaluation dataset. The results show that LP-RDFT provides effective recall and repeatability at a substantial reduction in computational cost and memory requirements compared to other descriptors.

Chapter 2 reviews the existing state of the art for local appearance description in images. Chapter 3 describes how the proposed method combines a Laplacian

of Gaussian with radial Fourier Transform to provide a local descriptor. Chapter 4 describes experiments with robustness and discusses the results. Chapter 5 summarizes our conclusions from this work.

2 Local descriptors

Since its introduction in 1999, the Scale Invariant Feature Transform (SIFT) descriptor [12] has remained the reference in local image description. The SIFT descriptor uses local maxima in the Laplacian pyramid to determine a reference scale for local appearance, and then describes local appearance at this scale using local histograms of the orientation of image derivatives calculated within a grid of small windows. Despite many challengers, SIFT, and its numerous variations continues to dominate the state of the art. In [14] it is shown that SIFT and similar descriptors performed the best. Recently, SURF [2] has been shown to provide results that are similar to SIFT with a reduced computational cost. A number of other local descriptors have recently been proposed, including ORB [16], BRISK [10], FREAK [1], BRIEF [5] and NSD [4]. None-the-less, SIFT remains the reference for image description due to its repeatability and recall.

As embedded computing and mobile computing applications for computer vision increase in popularity, memory requirements have emerged as an important issue. In response, many researchers have investigated the use of binary descriptors [4, 5, 10, 16]. As an alternative, we explore a compact descriptor with a variable vector length that can be adapted to meet the requirements of individual problems.

3 Creating the proposed descriptor

In order to create a robust multiscale descriptor, we investigated the use of two transformations with known invariant properties: the Laplacian of Gaussian Pyramid and the 2D Fourier Transform (figures 1 and 2). Both of these transformations are known to provide useful descriptions of local appearance. Our objective is to combine these two transformation to obtain a variable size image descriptor that preserves their invariant properties while improving discriminability.

A Gaussian pyramid is computed by convolving an image with Gaussian low pass filters with variances taken from an exponential series, such as 2^k [6]. Resampling each low pass image at a sample distance that is proportional to the standard deviation results in a set of images of exponentially decreasing size and identical impulse response. A Laplacian of Gaussian pyramid (or DoG pyramid) is created by subtracting each sampled low pass image from the next larger image in the pyramid. Each sample in the Laplacian Pyramid contains the value of the Laplacian (2nd derivative) of the image at a particular scale (variance) and position. Local maxima in the Laplacian of Gaussian can be used as keypoints for position and scale, as with the SIFT descriptor.

A Laplacian Profile (LP) is the vector of laplacian samples over an exponential set of scales at a specific image position. This vector is a co–variant with scale [11, 7]. Changes in image scale due to optical zoom or changes in distance result in an exact translation of the Laplacian Profile in the scale direction. Local extrema in the Laplacian profile can be used to determine the characteristic scale for image description.

We use the Half-Octave Gaussian pyramid algorithm [6, 17] to produce a Laplacian pyramid. The scaled images are convolved with a Gaussian filter $G(x, y, 2^k)$ for integer k and resampled with a sample distance of $s_k = 2^{(k-1)/2}$, keeping impulse response identical for each level. We then construct an LP vector for $p(x, y)$ collecting the Laplacian of Gaussian values for sample $p(x, y)$ from the k adjacent levels in the Laplacian pyramid. Although sampling results in some loss of invariance, this pyramid algorithm is highly efficient for computing LPs.

The LP can be computed at any image position. As with the Laplacian, the LP is also invariant to rotation. The length of the LP is variable and can be calculated on any height on the pyramid of scaled images.For example, an LP vector of length three can be collected in a six level pyramid at levels two-three-four and levels three-four-five. The top pyramid levels can be discarded because they are produced with an impulse response that is larger than the original image.

To improve the discriminative power, we use a Radial Discrete Fourier Transform (RDFT) around each sample of the Laplacian profile. In [13], the neighborhoods were either the 4 closest neighbors around each LP value, sampled on the Laplacian image pyramid (easily calculated by the Gaussian pyramid), or a disk of samples around each LP value, sampled on the Gaussian pyramid. We extended the approach by using 8 neighbors sampled on the Gaussian pyramid and calculate an 1D DFT linearly on them. This is an important extension from [13] as it works well with local description while keeping a small descriptor vector length. From the 8 neighbors x_0,x_1, ..., x_7, we take 8 Fourier coefficients X_0,X_1, ..., X_7. We keep the absolute value (magnitude) of X_0, the sign of X_4 and the magnitudes of X_1,X_2 and X_3. The absolute value of X_0 and the sign of X_4 are a measure of the sum of intensities of the 8 neighbors. The magnitudes of X_1,X_2 and X_3 provide frequency information. X_5,X_6 and X_7 provide similar information and are therefore discarded. Phase coefficients may be discarded or used to determine a characteristic orientation for the neighborhood.

In [13], an RDFT is computed around every element of an LP vector. This parameter can also be made variable to provide additional discriminative power. We concatenate the RDFTs to form a single description vector that is then normalized with the L_2-norm. The LP vector length, the LP elements around where the RDFT is computed and the radius where the 8 neighbors are taken, are all variable and can be selected according to the needs of an application.

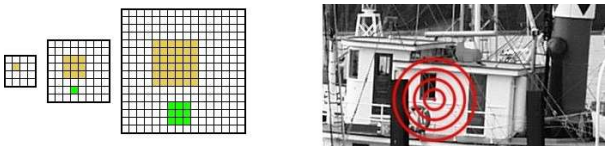

Fig. 1: A small neighborhood at a scale corresponds to a larger one at a lower scale. So, a multiscale vector captures information of increasingly larger areas on the image.

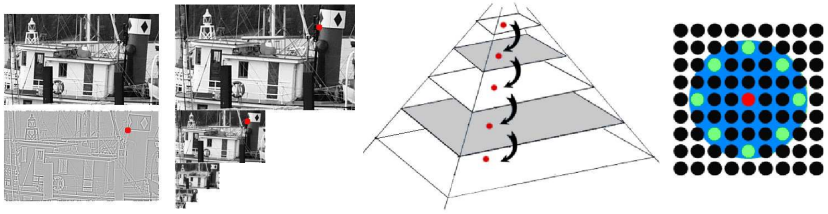

Fig. 2: **Left:** A Laplacian of Gaussian is easily computed on the levels of a pyramid of scaled images as a weighted difference of adjacent pixels. **Center:** Collecting a LP vector on the Gaussian pyramid. **Right:** Collection of 8 pixels around an LP element on a pyramid level for computing the RDFT. There is no limitation, concerning around which or how many of the elements of an LP, that the RDFT can be computed.

4 Experiments

4.1 Textured images

We evaluate LP-RDFT against the well established local descriptors ORB, BRISK, FREAK, SIFT, SURF, BRIEF and the newly proposed NSD in its two forms, the Seed of Life (SOL) and the Binary Seed of Life (BinSOL). We use the Affine Covariant Features benchmark dataset [14] for the experiments. The test is keypoint matching between images. The testing protocol is proposed on the dataset web site: compare the first image in each case folder with the rest of the images in the same case folder. The images are highly textured, and thus result in variations in the image signal that provide rich information about content.

The measures we use for the evaluation are repeatability [15, 9] and recall against 1 - precision [14]. Repeatability shows how good a method is at finding correct matches. Recall against 1-precision plots show how important is the quantity of correct matches found by a method considering the quantity of false matches. We use the Euclidean distance for matching the descriptor vectors. For all the descriptors except NSD, we use the OpenCV library [3]. For NSD we use the source code provided by the authors. The parameters for the descriptors are kept at their default values, trusting that their authors and developers have made the best choices.

For LP-RDFT, after some experimentation, we concluded that the best performance was given by creating LP vectors of length seven (seven exploited

pyramid levels). The same experiments showed that for the computation of the RDFT, the best choice is to collect samples at a radius of five pixels around the LP coordinates for the highest four of the seven pyramid levels. These choices create a descriptor vector of 27 real valued elements. The smallest descriptors in the competing test set are the binary descriptors ORB and BRIEF with 256 binary elements (bits) stored in 32 bytes (OpenCV implementation).

For each descriptor we use the keypoint detector proposed by its authors and developers. For LP-RDFT, we collect keypoints using DoG to compute the Laplacian pyramid. The results of keypoint matching is shown in four figures, numbers 3 to 6. Each figure has two pairs of graphs, each pair corresponding to a particular case folder of the dataset. For each graph pair, one graph shows the repeatability measure for each couple of compared images (image 1 to another image in the same case folder, characterized by an index number) and one graph shows the recall against 1 - precision.

As we can see from the figure 3, for the cases of increasing blur ("bikes"and "trees"), LP-RDFT works very well compared to the state of the art, with very competitive rates for both correct and false matches as depicted by the repeatability and recall against 1 - precision plots. For the right pair of graphs in figure 3 for the case folder "trees", LP-RDFT outperforms the other methods. In both cases, LP-RDFT outperforms SIFT. The right pair of figure 6 for the case of increasing JPEG compression ("ubc") shows that LP-RDFT has a very competitive performance, with high readability and few false matches. For high values of JPEG compression, LP-RDFT outperforms all other descriptors. These results show that LP-RDFT performs well when information is lost due to bad resolution or image compression, regardless of its very small vector length.

Figure 4, for viewpoint changes ("graf"and "wall"), shows that the proposed method has a mediocre repeatability compared to the state of the art. It also provides comparable amount of false matches to the other descriptors as can be concluded by the recall against 1 - precision plots. Figure 5 for zoom and rotation changes ("bark"and "boat"), shows also lower performance than most of the state of the art. Again LP-RDFT provides comparable number of false matches. In figure 6, the left pair for the case of decreasing light ("leuven"), shows that the proposed method has a lower performance compared to the state of the art.

LP-RDFT performs very competitively to the state of the art with a very small vector of only 27 elements, especially when the higher frequencies are lost. The method works very well for blur and JPEG compression, which are relevant to image scaling, but it is weaker at viewpoint changes and light variations.

4.2 Textureless images

Textureless images are more difficult to discriminate because they exhibit fewer visual features, smooth edges and large homogeneous areas. In order to make a more general testing, we collected a set of images (figure 7) from the MIR-FLICKR Retrieval Evaluation dataset [8] with these characteristics.

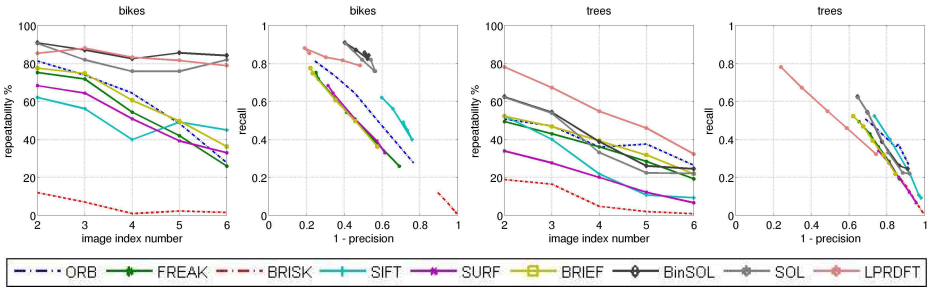

Fig. 3: Affine Covariant Features dataset. **Two left graphs:** Blur ("bikes"). **Two right graphs:** Blur ("trees"). Keypoint matching of image 1 to the rest.

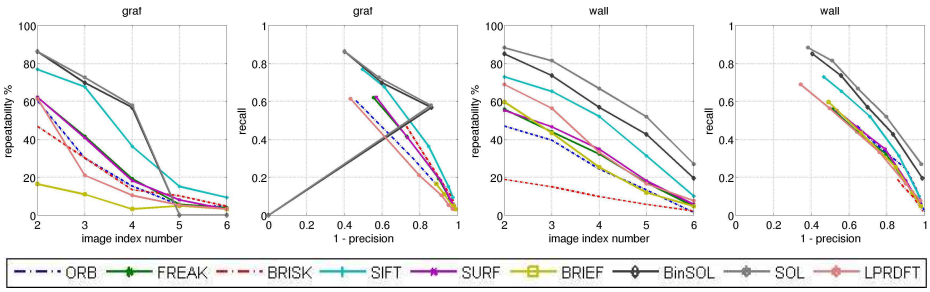

Fig. 4: Affine Covariant Features dataset. **Two left graphs:** Viewpoint ("graf"). **Two right graphs:** Viewpoint ("wall"). Keypoint matching of image 1 to the rest.

The tests include rotation of the images from 0° to 180° every 30° and scaling of the images from four times bigger to four times smaller with scale factor $\sqrt{2}$ (figure 8). The same measures are used for evaluation. All compared descriptors are used in the same way as before. Concerning the proposed descriptor, we use another shorter version of LP-RDFT due to the size of the images. Also, the radius, where we collect the samples for the RDFT, is a little bigger; six pixels instead of five. Again, the parameters are chosen after exhaustive tests. We use two element long LP vectors, so two exploited levels on the image pyramid, and the Fourier information from only the highest of the two used pyramid levels. The final vector is thus particularly small with only 7 elements!

From the plots we see that none of the methods give high performance, with low recall and repeatability. The low recall against 1-precision measures results from the small number of keypoints detected and used for matching. Surprisingly, for all descriptors except LP-RDFT, the matching performance on the same image (original image to its self) is very low. This can be explained by the lack of meaningful signal information in this images that causes many of the descriptor features to look alike and cause false matches. The proposed method performs almost perfectly for matching on the original image, which shows that it can handle low quality information. Its performance though deteriorates with

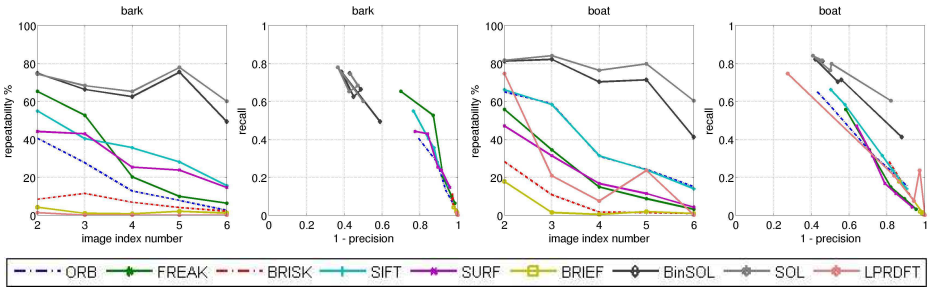

Fig. 5: Affine Covariant Features dataset. **Two left graphs:** Zoom + rotation ("bark"). **Two right graphs:** Zoom + rotation ("boat"). Keypoint matching of image 1 to the rest.

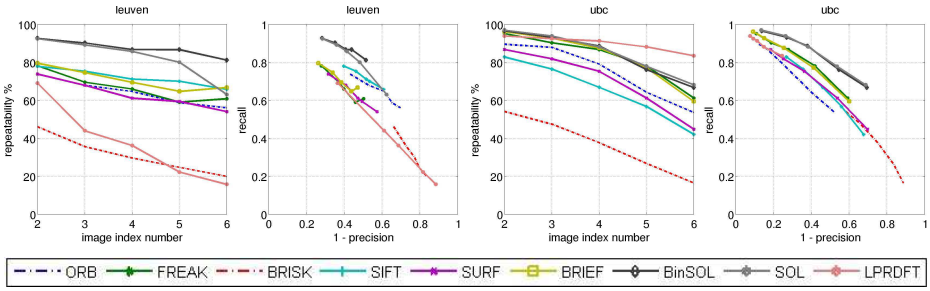

Fig. 6: Affine Covariant Features dataset. **Two left graphs:** Light ("leuven"). **Two right graphs:** JPEG compression ("ubc"). Keypoint matching of image 1 to the rest.

rotation and scaling. For small rotations and scale changes the repeatability of LP-RDFT is the best among all descriptors.

5 Conclusions

In this work, we proposed a method of image description by combining two transformations with interesting invariant properties with the goal of creating a very compact and robust image descriptor. The resulting LP-RDFT is a multiscale descriptor with variable vector size that can become very small if needed while remaining discriminative enough. The experimental results on keypoint matching for textured images showed that LP-RDFT works efficiently having a very small vector length. LP-RDFT outperforms the state of the art for scale changes and image changes relevant to scaling, like increasing blur and JPEG compression. Tests on textureless images showed that LP-RDFT beats the state of the art for small values of rotation and scaling but its performances deteriorates for larger values. The most important fact is that the vector size of LP-RDFT for the textureless images tests is particularly tiny, with only 7 elements. LP-RDFT is a proposition for very compact image description and especially suitable for

Fig. 7: The textureless images taken from MIRFLICKR Retrieval Evaluation dataset.

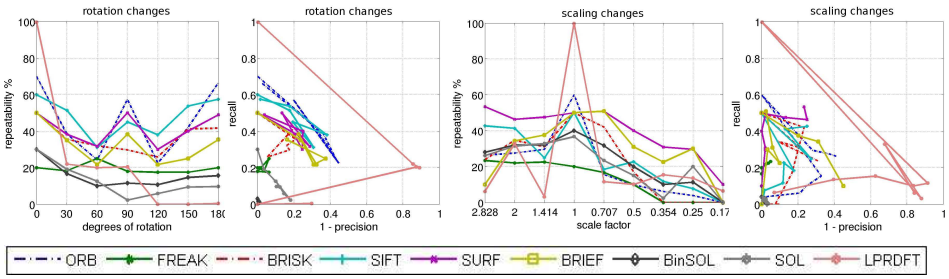

Fig. 8: Textureless images from MIRFLICKR Retrieval Evaluation dataset. **Two left graphs:** Rotation tests. **Two right graphs:** Scaling tests.

cases where a large amount of information is unnecessary and large vector sizes can be a problem.

References

1. Alahi, A., Ortiz, R., Vandergheynst, P.: FREAK: Fast Retina Keypoint. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 510–517 (June 2012)
2. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-Up Robust Features (SURF). Comput. Vis. Image Underst. 110(3), 346–359 (June 2008), `http://dx.doi.org/10.1016/j.cviu.2007.09.014`
3. Bradski, G.: The OpenCV Library. Dr. Dobb's Journal of Software Tools (2000)
4. Byrne, J., Shi, J.: Nested Shape Descriptors. In: 2013 IEEE International Conference on Computer Vision (ICCV). pp. 1201–1208 (December 2013)
5. Calonder, M., Lepetit, V., Ozuysal, M., Trzcinski, T., Strecha, C., Fua, P.: BRIEF: Computing a Local Binary Descriptor Very Fast. IEEE Trans. on Pattern Analysis and Machine Intelligence 34(7), 1281–1298 (2012)
6. Crowley, J.L., Stern, R.M.: Fast Computation of the Difference of Low-Pass Transform. IEEE Trans. on Pattern Analysis and Machine Intelligence PAMI-6(2), 212–222 (1984)
7. Hall, D., Verdiere, V., Crowley, J.L.: Object Recognition Using Coloured Receptive Fields. In: Computer Vision - ECCV 2000, Lecture Notes in Computer Science, vol. 1842, pp. 164–177. Springer Berlin Heidelberg (2000), `http://dx.doi.org/10.1007/3-540-45054-8_11`
8. Huiskes, M.J., Lew, M.S.: The MIR Flickr Retrieval Evaluation. In: MIR. ACM, New York, NY, USA (2008)
9. Juan, L., Gwon, O.: A Comparison of SIFT, PCA-SIFT and SURF. International Journal of Image Processing (IJIP) 3(4), 143–152 (2009)

10. Leutenegger, S., Chli, M., Siegwart, R.: BRISK: Binary Robust invariant scalable keypoints. In: 2011 IEEE International Conference on Computer Vision (ICCV). pp. 2548–2555 (November 2011)
11. Lindeberg, T.: On the axiomatic foundations of linear scale-space: Combining semi-group structure with causality vs. scale invariance, Technical report, Department of Numerical Analysis and Computing Science, Royal Institute of Technology, S-100 44 Stockholm, Sweden, August 1994. (ISRN KTH NA/P–94/20–SE). Revised version published as Chapter 6 in J. Sporring, M. Nielsen, L. Florack, and P. Johansen (eds.) *Gaussian Scale-Space Theory: Proc. PhD School on Scale-Space Theory*, (Copenhagen, Denmark, May 1996), Kluwer Academic Publishers, 75–98, (1997)
12. Lowe, D.: Object recognition from local scale-invariant features. In: The Seventh IEEE International Conference on Computer Vision. vol. 2, pp. 1150 –1157 (1999)
13. Mavridou, E., Hoàng, M.D., Crowley, J.L., Lux, A.: Scale normalized radial fourier transform as a robust image descriptor. In: ICPR (2014 In press)
14. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. on Pattern Analysis and Machine Intelligence 27(10), 1615–1630 (2005)
15. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Van Gool, L.: A Comparison of Affine Region Detectors. International Journal of Computer Vision 65(1-2), 43–72 (November 2005), `http://dx.doi.org/10.1007/s11263-005-3848-x`
16. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: An Efficient Alternative to SIFT or SURF. In: Computer Vision (ICCV), 2011 IEEE International Conference on. pp. 2564–2571 (November 2011)
17. Ruiz-Hernandez, J.A., Lux, A., Crowley, J.L.: Face detection by cascade of Gaussian derivates classifiers calculated with a Half-Octave Pyramid. In: 8th IEEE International Conference on Automatic Face Gesture Recognition, 2008. FG '08. pp. 1 –6 (September 2008)