

A NOVEL RANK ORDER LOG FILTER FOR INTEREST POINT DETECTION

Zhenwei Miao and Xudong Jiang

School of Electrical and Electronic Engineering,
Nanyang Technological University, Singapore 639798

{miao001ei, exdjiang}@ntu.edu.sg

ABSTRACT

This paper proposes a novel non-linear filter, named rank order LoG (ROLG) filter, and a new interest point detector, named ROLG detector. The ROLG filter is a weighted rank order filter. It is used to detect image structures whose significant majority of pixels are brighter (or darker) than the significant majority of pixels in their corresponding surroundings. The ROLG detector is built on this filter. Compared to linear filter based detectors, the proposed rank order filter based detector is more robust to abrupt variations of images. Experiments on the benchmark databases demonstrate that the ROLG detector achieves superior performance compared to four state-of-the-art detectors. Evaluation experiments are also conducted on face recognition. The results further demonstrate that the ROLG detector has better performance compared to other detectors.

Index Terms— Interest point detection, image matching, weighted rank order filter, repeatability, face recognition.

1. INTRODUCTION

As a powerful tool for computer vision, interest point detection has drawn great attentions in the last two decades [1, 2], and has been successfully used in many applications, e.g. [3, 4]. Interest point detectors can be roughly classified into three categories: corner-based detectors, blob-based detectors and region detectors.

Corners correspond to points in the 2D images with high curvature [2]. Harris corner detector [5] uses the second moment matrix to detect the local image structures, where the intensity changes in two orthogonal directions are both large. Harris-Laplace/Affine detectors [6] are proposed to extend Harris corner detector into multiple scales and be covariant with scale and affine changes. SUSAN detector [7] defines corners as the smallest USAN (univalue segment assimilating nucleus) points, which are dissimilar from a majority of pixels within their surrounding regions.

Blobs refer to bright regions on dark backgrounds or vice versa [8]. Hessian detector [9] employs the Hessian matrix to detect blobs in a single scale. Hessian-Laplace/Affine detectors [6] are developed to detect blobs in multiple scales. SIFT detector [10] is proposed to speed up the detecting process by employing the difference of Gaussian (DoG) filter. The DoG filter approximates the normalized Laplacian of Gaussian (LoG) filter.

Region detectors extract regions with similar image structures and properties [2]. Intensity-based region detector is proposed in [11] to detect affine invariant regions based on the intensity variation along rays emanating from the intensity extremum. Maximally stable extremal regions (MSER) are extracted in [12] to detect affine invariant regions.

Our work in this paper is inspired by the SIFT detector, and intends to solve the problems of this detector. As one of the most

popular detectors, the SIFT detector [10] employs the DoG filter to generate the blob map. The DoG filter is a linear filter. Its response is easily effected by the strong and abrupt structures near the structure to be detected. Moreover, a set of unstable spurious points may be detected around the structures due to the second order derivative nature of the SIFT detector.

To alleviate the problems of the SIFT detector, we propose a novel weighted rank order filter with weights proportional to the coefficients of the LoG filter. This non-linear filter is named rank order Laplacian of Gaussian (ROLG) filter. It is used to detect image structures whose majority of pixels are brighter (or darker) than the majority of pixels in their corresponding surroundings. The proposed new interest point detector is built on the ROLG filter to detect interest points in multiple scales.

2. THE PROPOSED ROLG FILTER

As a necessary preliminary of the study, we discuss the properties of the LoG filter and the weighted rank order filter in Section 2.1 and 2.2, respectively.

2.1. LoG Filter

The LoG filter (Fig. 1(a)) has been applied in interest point detectors [6, 9, 10, 13]. It is defined by

$$w(x, y, \sigma) = -\frac{1}{\pi\sigma^4} \left[1 - \frac{x^2 + y^2}{2\sigma^2} \right] e^{-\frac{x^2 + y^2}{2\sigma^2}}, \quad (1)$$

where σ is the standard deviation of the Gaussian function and also named scale factor. The output of the LoG filter at (x, y) is given by

$$r(x, y, \sigma) = \sum_{(m,n) \in S_1} w_+(m, n, \sigma) I(x-m, y-n) - \sum_{(m,n) \in S_2} w_-(m, n, \sigma) I(x-m, y-n). \quad (2)$$

S_1 and S_2 (as shown in Fig. 1(b)) are the two parts of the filter mask. S_1 contains the positive weights and S_2 contains all the negative weights. w_+ and w_- are the absolute values of the weights in S_1 and S_2 , respectively. The output of the LoG filter is the difference between the weighted average of pixels in S_1 and S_2 .

This filter is ineffective to deal with the sparse but strong noise, such as the pepper&salt noise. Even a small portion of pixels may greatly affect the output adversely if their grey values largely deviate from those of the image structure to be detected. However, a small portion of pixels have almost no effect on the output of the rank order filter even if their grey values are extremely high or low. This motivates us to design a weighted rank order filter with weights proportional to those of the LoG filter for interest point detection.

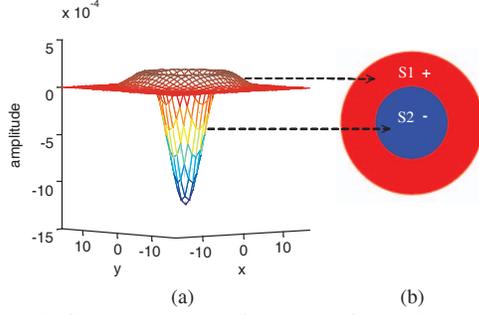


Fig. 1. LoG filter. (a) shape of the LoG filter. (b) two parts of the LoG filter, S_1 corresponds to the surrounding ring containing the positive weights, S_2 corresponds to the inner circular disk containing all the negative weights.

2.2. Weighted Rank Order Filter

The output of the weighted rank order filter [14] is defined as follows. Assuming the weights for the input series $\mathbf{x} = \{x_1, x_2, \dots, x_q\}$ are $\mathbf{w} = \{w_1, w_2, \dots, w_q\}$. For the ascending sorted $\tilde{\mathbf{x}} = \{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_q\}$, their corresponding weights are rearranged as $\tilde{\mathbf{w}} = \{\tilde{w}_1, \tilde{w}_2, \dots, \tilde{w}_q\}$. The output of the weighted rank order filter with rank r_w , $r_w \in \{1, 2, \dots, \sum_{j=1}^q w_j\}$ is represented as

$$\begin{aligned} y_{r_w} &= \text{rank}_{r_w} \{w_1 \diamond x_1, w_2 \diamond x_2, \dots, w_q \diamond x_q\} \\ &= \text{rank}_{r_w} \{\tilde{w}_1 \diamond \tilde{x}_1, \tilde{w}_2 \diamond \tilde{x}_2, \dots, \tilde{w}_q \diamond \tilde{x}_q\}, \end{aligned} \quad (3)$$

in which \diamond is the replication operator defined by

$$w_i \diamond x = \underbrace{x, x, \dots, x}_{w_i \text{ times}}. \quad (4)$$

By defining a cumulative sum of the sorted weights as

$$c_i = \sum_{j=1}^i \tilde{w}_j, \quad (5)$$

for $i \in \{1, 2, \dots, q\}$ and $c_0 = 0$, the output of the weighted rank order filter is given by

$$y_{r_w} = \tilde{x}_{i_o}, \{i_o : c_{i_o-1} < r_w \leq c_{i_o}\}. \quad (6)$$

2.3. The Proposed ROLG Filter

One direct way to employ the weighted rank order filter for detecting interest points is to replace the weighted average in (2) by the weighted median, as

$$\begin{aligned} r_{wm}(x, y, \sigma) &= \text{median}_{(m,n) \in S_1} (\hat{w}_+(m, n, \sigma) \diamond I(x-m, y-n)) - \\ &\quad \text{median}_{(m,n) \in S_2} (\hat{w}_-(m, n, \sigma) \diamond I(x-m, y-n)), \end{aligned} \quad (7)$$

where $\hat{w}_+(m, n, \sigma) = w_+(m, n, \sigma) / \sum w_+$ and $\hat{w}_-(m, n, \sigma) = w_-(m, n, \sigma) / \sum w_-$. With these weighting coefficients, more important pixel has higher influence to the filter output. The difference between the two weighted median filters (7) has similar role to the LoG filter (2) and, hence, can be used to detect interest points.

However, when noise exists, filter (7) produces very strong response on an edge if one median filter captures one side of the edge while the other median filter happens to capture the other side of the edge. Additional rules are needed to enhance the robustness of the

detector. As we know, median filter is a special case of rank order filter, as median is equal to rank 0.5. Firstly, the weighted median filters in (7) are replaced by the weighted rank order filters. Then, (7) is reformulated as

$$\begin{aligned} r_{wr}(x, y, \lambda_1, \lambda_2) &= \text{rank}_{\lambda_1} (\hat{w}_+(m, n) \diamond I(x-m, y-n)) - \\ &\quad \text{rank}_{\lambda_2} (\hat{w}_-(m, n) \diamond I(x-m, y-n)), \end{aligned} \quad (8)$$

where λ_1 and λ_2 are the rank factors for the two weighted rank order filters.

In order to suppress the edge response, we require a significant majority of pixels ($> 50\%$, e.g. 60%) in the surrounding ring larger than a significant majority of pixels ($> 50\%$, e.g. 60%) in the inner disk, or a significant majority of pixels in the surrounding ring smaller than a significant majority of pixels in the inner disk. Otherwise, the outputs are set to zero to suppress noises and edges. It is not too difficult to see that this idea can be realized by introducing a positive nonzero offset parameter δ and let $\lambda_1 = 0.5 - \delta$, $\lambda_2 = 0.5 + \delta$ if the resulting output (8) is still positive, and let $\lambda_1 = 0.5 + \delta$, $\lambda_2 = 0.5 - \delta$ if the resulting output (8) is still negative and otherwise set the output (8) zero. With this idea, the proposed ROLG filter is defined by

$$\begin{aligned} r_{ROLG}(x, y, \sigma, \delta) &= \\ &\begin{cases} \mathcal{P} = r_{wr}(x, y, \sigma, 0.5 - \delta, 0.5 + \delta), & \text{if } \mathcal{P} > 0 \\ \mathcal{N} = r_{wr}(x, y, \sigma, 0.5 + \delta, 0.5 - \delta), & \text{if } \mathcal{N} < 0 \\ 0, & \text{otherwise} \end{cases} \end{aligned} \quad (9)$$

3. INTEREST POINT DETECTION BY ROLG FILTER

The response of the ROLG filter on a ridge is large if the scale of the ROLG filter is approximately the same as the width of the ridge. Points detected on ridges are sensitive to noise. Such kind of unstable points are removed as done in [10]. Interest point detection in multiple scales is an important issue in vision applications. In our implementation, we employ a straightforward method by detecting interest points in each scale as done in [15].

The proposed algorithm for the ROLG detector is summarized as follow:

- 1: Initialize the ROLG filter by setting the offset parameter δ and the scale parameter σ .
- 2: Generate the corner/blob map by filtering the input image with the ROLG filter.
- 3: Detect peaks on the corner/blob map, and remove peaks which are on ridges. Remaining peaks are the interest points in this scale.
- 4: Update the ROLG filter by a larger scale σ , and go back to step 2 to detect interest points in a new scale until the maximal scale is reached.

4. EXPERIMENTS

We evaluate the proposed ROLG detector in three experiments: 1) visual comparison, 2) quantitative evaluation using the repeatability and the matching score [1], and 3) quantitative evaluation in the application of face recognition. In experiments 2 and 3, the SIFT descriptor [10] is used to describe the interest points for all detectors.

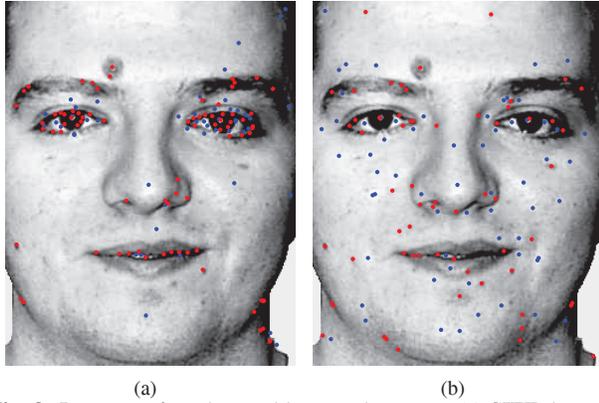


Fig. 2. Interest points detected by two detectors. (a) SIFT detector, (b) ROLG detector. Local maxima are indicated in blue, and local minima are indicated in red.

4.1. Visual Comparison

Fig. 2 gives a visual comparison between the SIFT detector and the ROLG detector on a face image. The parameters are chosen so that the same number of points are detected by the both detectors. We see that SIFT detects many spurious points around the eyeballs while misses a lot of points in other areas. The ROLG detector alleviates these problems.

4.2. Repeatability and Discrimination Tests

Two metrics, one is repeatability and the other is matching score as defined in [1], are used. The “repeat” and “match” of two detected regions in a given image pair are defined as follow [1]. Two regions are repeated if their overlap error is below a certain threshold (in our experiment, the threshold is 40%). Two regions are matched if 1) they are repeated, and 2) their descriptors are the nearest neighbor in the descriptor space. The repeatability score for a given image pair is the ratio between the repeated regions and the larger number of detected regions in the image pair within the common area and the common scales. The matching score for a given image pair is the ratio between the number of correct matches and the larger number of detected regions in the image pair within the common area and the common scales.

Four data set, corresponding to scale change for textured scene, viewpoint change for textured scene, blur for textured scene, and illumination change, respectively, are chosen from the database [1]. Each data set consists of 6 images with 5 homographies between the first image and the other five images. In each data set, the first image is used as the reference image and remaining 5 are used as the test images.

In all the experiments here, interest points are detected on the downsampled images. The offset parameter δ is set to 0.1 to suppress noises and edges. Interest points are detected in 12 scales: $\{\sigma_n\}_{n=1,2,\dots,12} = \{1.6 \times 2^{1/3}, 1.6 \times 2^{2/3}, 3.2, \dots, 1.6 \times 2^4\}$. Instead of continuously increasing the ROLG mask size, the 12 scales are divided into 4 octaves by downsampling the previous octave. Each octave contains 3 scales $\{\sigma_{no}\}_{no=1,2,3} = \{1.6 \times 2^{1/3}, 1.6 \times 2^{2/3}, 3.2\}$.

Four detectors, the MSER detector [12], the Harris-affine (HR-A) detector [6], the Hessian-affine (HS-A) detector [6] and the SIFT detector [10], are compared with the ROLG detector. The default parameters given by the authors are used for each detector. Exper-

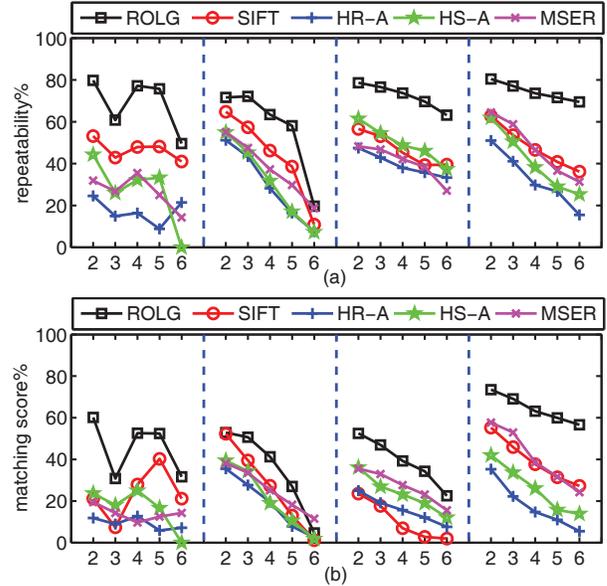


Fig. 3. Repeatability (a) and matching score (b) on the four data set. In each column, horizontal axis represents the image index in the corresponding data set. From left to right of (a) and (b) are the results on the scale change for textured sequence, the viewpoint change for textured sequence, the blurring for textured sequence, and the illumination change sequence, respectively.

iment results shown in Fig. 3 demonstrate that the ROLG detector outperforms the other 4 detectors in almost all cases.

4.3. Application to Face Recognition

AR [16], ORL [17], Georgia Tech (GT) [18], FERET [19] and Labeled Faces in the Wild (LFW) [20] databases are chosen to test the discriminative power of the interest points in face recognition.

AR database: Color images are converted to gray images and normalized into the size of 60×85 . 75 subjects with 14 nonoccluded images per person are selected. The first 7 images of all subjects are chosen as gallery set, and the remaining 7 images as probe set.

ORL database: Images are normalized into the size of 50×57 . The first 5 images of all 40 subjects are chosen as gallery set, and the remaining 5 images as probe set.

GT database: Color images are converted to gray images and normalized into the size of 60×80 . The first 8 images of all 50 subjects are chosen as gallery set, and the remaining 7 images as probe set.

FERET database: Images are cropped into the size of 60×80 . 1194 subjects with 2 images per person are selected. The first 1 image of all subjects is chosen as gallery set, and the remaining 1 image as probe set.

LFW database: Color images are converted to gray images and cropped into the size of 64×64 . 134 subjects with 10 images per person are selected. The first 5 images of all subjects are chosen as gallery set, and the remaining 5 images as probe set.

Face recognition is an active research topic [21, 22, 23, 24] and some work has been done to apply the SIFT detector and descriptor in face recognition [25]. In this experiment, we compare the ROLG detector with 4 state-of-the-art detectors, the SIFT detector [10], the SURF detector [13], the Harris-affine (HR-A) detector [6], and the Hessian-affine (HS-A) detector [6]. When using the default parame-

ters given by the authors, all detectors detect too few points and lead to very poor performance. Thus, we decrease the contrast threshold and find that zero is the best for all detectors. Hence, the thresholds used to remove the low contrast interest points are set to zero for all detectors. The matching procedures described in [10] are employed in this experiment.

The recognition error rates are shown in Table 1. For the AR, ORL, GT, and FERET databases, the ROLG detector significantly outperforms the other 4 detectors. For the LFW database, significant variations of expression, pose, illumination and occlusion exist. These variations lead to poor performance for all detectors here. However, the ROLG detector still outperforms the other 4 detectors.

Table 1. Recognition error rates on AR, ORL, GT, FERET and LFW databases.

	AR	ORL	GT	FERET	LFW
ROLG	1.7%	3.5%	8.9%	1.8%	63.6%
SIFT	5.7%	10.0%	16.0%	10.1%	72.4%
SURF	7.4%	21.5%	15.4%	10.4%	79.1%
HS-A	11.4%	20.0%	26.0%	14.7%	84.0%
HR-A	25.5%	33.5%	52.6%	50.3%	89.6%

5. CONCLUSION

A novel non-linear filter named rank order Laplacian of Gaussian (ROLG) filter is proposed, and a new interest point detector named ROLG detector is designed in this paper. The ROLG filter is a weighted rank order filter. Compared to the SIFT detector, the ROLG detector detects less spurious, unstable points and is more robust to abrupt variations of images. Experiment results demonstrate that its performance is better compared to 4 state-of-the-art detectors in terms of the repeatability and discrimination of the interest points. The application to the face recognition on five databases further verifies the superiority of the ROLG detector.

6. REFERENCES

- [1] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," *Int. J. Computer Vision*, vol. 65, no. 1-2, pp. 43–72, 2005.
- [2] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: a survey," *Fundations and Trends in Computer Graphics and Vision*, vol. 3, no. 3, pp. 177–280, 2008.
- [3] M. Brown and D.G. Lowe, "Recognising panoramas," in *Proc. Ninth Int. Conf. Computer Vision*, 2003, pp. 1218–1225.
- [4] C. Schmid and R. Mohr, "Local grayvalue invariants for image retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 5, pp. 530–535, 1997.
- [5] C. Harris and M. Stephens, "A combined corner and edge detector," in *Alvey Vision Conference*, 1988, pp. 147–151.
- [6] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *Int. J. Computer Vision*, vol. 60, no. 1, pp. 63–86, 2004.
- [7] S.M. Smith and J.M. Brady, "SUSAN - a new approach to low level image processing," *Int. J. Computer Vision*, vol. 23, no. 1, pp. 45–78, 1997.
- [8] T. Lindeberg, "Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention," *Int. J. Computer Vision*, vol. 11, no. 3, pp. 283–318, 1993.
- [9] P. R. Beaudet, "Rotationally invariant image operators," in *Proc. Int. Conf. Pattern Recognition*, 1978, pp. 579–583.
- [10] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [11] T. Tuytelaars and L. Van Gool, "Matching widely separated views based on affine invariant regions," *Int. J. Computer Vision*, vol. 59, no. 1, pp. 61–85, 2004.
- [12] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and Vision Computing*, vol. 22, no. 10, pp. 761–767, 2004.
- [13] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.
- [14] G.R. Arce, *Nonlinear Signal Processing: A Statistical Approach*, New York: Wiley, 2004.
- [15] W.T. Lee and H.T. Chen, "Histogram-based interest point detectors," in *Proc. Conf. Computer Vision and Pattern Recognition*, 2009, pp. 1590–1596.
- [16] A.M. Martinez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 748–763, 2002.
- [17] F. Samaria and A. Harter, "Parameterisation of a stochastic model for human face identification," in *Second IEEE Workshop Applications of Computer Vision*, 1994, pp. 138–142.
- [18] "Georgia Tech Face Database," http://www.anefian.com/face_reco.htm, 2007.
- [19] P.J. Phillips, H. Moon, S.A. Rizvi, and P.J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.
- [20] G.B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Tech. Rep. 07-49, University of Massachusetts, Amherst, 2007.
- [21] X.D. Jiang, B. Mandal, and A. Kot, "Eigenfeature regularization and extraction in face recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 383–394, 2008.
- [22] X.D. Jiang, "Asymmetric principal component and discriminant analyses for pattern classification," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 5, pp. 931–937, 2009.
- [23] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.
- [24] X.D. Jiang, "Linear subspace learning-based dimensionality reduction," *IEEE Signal Processing Magazine*, vol. 28, no. 2, pp. 16–26, 2011.
- [25] C. Geng and X.D. Jiang, "Face recognition based on the multi-scale local image structures," *Pattern Recognition*, vol. 44, no. 10-11, pp. 2565–2575, 2011.