Fast LoG SIFT Keypoint Detector

Paras Maharjan, Lyle Vanfossan, Zhu Li
Dept. Computer Science & Engineering
University of Missouri-Kansas City
Kansas City, USA
{paras.maharjan, ltv7d8}@mail.umkc.edu, zhu.li@ieee.org

Jerry Jialie Shen
Dept. of Computer Science
City, University of London
London, UK
jerry.shen@city.ac.uk

2023 IEEE 25th International Workshop on Multimedia Signal Processing (MMSP) | 979-8-3503-3893-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/MMSP59012.2023.10337661

***Abstract*—Scale-invariant feature transform (SIFT) is a classical computer vision technique for scale-invariant keypoint detection and feature extraction. SIFT exhibits invariance to various transformations such as scale, rotation, noise, and illumination, making it applicable in a wide range of applications like object recognition, image matching and stitching, environment mapping, navigation, robotics, camera calibration, and more. A key contribution of SIFT is its utilization of the Difference-of-Gaussian (DoG) feature pyramid, which approximates the scale-space response of the Laplacian-of-Gaussian (LoG) filter. The DoG feature pyramid is computed by taking the separable Gaussian filtering and stacking the difference of Gaussian blurred images. In this paper, we propose a novel approach called "Fast LoG" filtering, which offers direct computation of the LoG filter to model the scale-space response solution. The "Fast LoG" filter is achieved by decomposing the LoG filter into two separable filters via SVD, and the scale-space response is computed by a direct polynomial fitting and differentiation, which is analytically more accurate. The polynomial fitting and differentiation only happen after the LoG peak strength thresholding, therefore the overall complexity is low compared with the DoG-based SIFT. The experimental results show that the keypoint generated by the Fast LoG method matches the SIFT keypoints, and per-pixel filtering complexity is lower.**

***Index Terms*—Keypoint detector, SIFT, Difference-of-Gaussian (DoG), Laplacian-of-Gaussian (LoG), Singular Value Decomposition (SVD)**

I. Introduction

Keypoint detectors play a crucial role in various computer vision applications, including image matching and stitching, object and scene recognition, Simultaneous Localization and Mapping (SLAM) for 3D reconstruction, camera calibration, etc. Over the years, techniques such as the Harris corners detector [1] and Scale-invariant Feature Transform (SIFT) [2], SURF [3], FAST [4], BRIEF [5], and many others [6, 7, 8, 9] have been widely employed in these fields. An effective keypoint representation should possess the ability to detect a concise and informative set of points that are relevant to the specific task at hand. It should also exhibit robustness against diverse transformations, such as scale changes, translations, noise, and variations in lighting conditions.

SIFT [2] is the most used keypoint detector and feature extractor due to its robustness in handling diverse image transformations. SIFT uses a Difference-of-Gaussian (DoG) approximation of the Laplacian-of-Gaussian (LoG) filter to generate the scale-space pyramid. Keypoints are then identified at locations where the DoG has extrema values, ie. maxima or minima, considering both the spatial coordinates in the image domain and the scale level within the pyramid structure. However, since DoG cannot capture the exact detail of the LoG filter, it may lead to potential loss of accuracy in keypoint detection. Hence in our work, we address this issue by utilizing a more precise LoG filter approximation using Singular Value Decomposition (SVD).

The extrema detected from the LoG filter response are used as a means of edge and blob detection, first by convolving an image with Gaussian filters and then applying the Laplacian function to the result. When done over several strengths of Gaussian scaling and then stacked into a pyramid structure, keypoint features can be derived through the mapping of maxima and minima of intensity ranges across an image. The resulting features are then usable in complex real-world scenarios where inconsistent object orientation, variable scaling, and low illumination are expected. However, the drawback of high computational cost, mainly due to the inseparable nature of the LoG filter, has limited its usability in practical applications, a more cost-effective DoG filter is used to approximate LoG through comparison of Gaussian blur at different scales, modifying computationally intensive Laplacian functions to simple subtraction of matrices. The improvements of DoG have led to its use in the Scale Invariant Feature Transform (SIFT) algorithm [2] as a means of scale-space extrema detection. Nevertheless, by employing singular value decomposition (SVD) on the LoG output, we found that there exist two non-zero singular values. This indicates that the LoG filter can be approximated using SVD separable filters, resulting in the generation of more precise keypoints.

In our research, we have introduced a novel approach for fast keypoint detection called the "Fast LoG" SIFT detector that utilizes an SVD separable LoG filter to construct a scale-space response pyramid. By fitting a polynomial, we directly detect extrema in the scale-space pyramid. Notably, the polynomial fitting and differentiation processes occur after the elimination of non-maximum responses in the LoG peak strength thresholding, resulting in an overall lower complexity when compared to the SIFT approach based on DoG. The generated keypoints exhibit a strong correlation with SIFT while demonstrating enhanced robustness and lower computational complexity compared to the DoG pyramid method from an off-the-shelf SIFT library called "VLFeat" [10].

Earlier keypoint detection methods like Harris-corner de-

Algorithm 1 Fast LoG SIFT

1. Load image I.
2. Compute 2-SVD approximation LoG scale-space response of I.
  i. Compute the scales $\sigma_i = a\sqrt{b^{i-1}}$
  ii. Compute scale invariant LoG response.
  $I_{LoG} = \sigma_i^2 LoG(I, \sigma_i)$
  Generate LoG scale-space pyramid.
3. Evaluate peak strength thresholding.
  i. Keypoint candidate $\leftarrow LoG(x, y, \sigma_i) \geq P_{th}$
4. Selecting the keypoint.
  i. Apply curve fitting f to the peak response for each keypoint candidate.
  ii. Keypoint $(K) \leftarrow abs(f'_{\sigma_{max}}) \leq G_{th}$

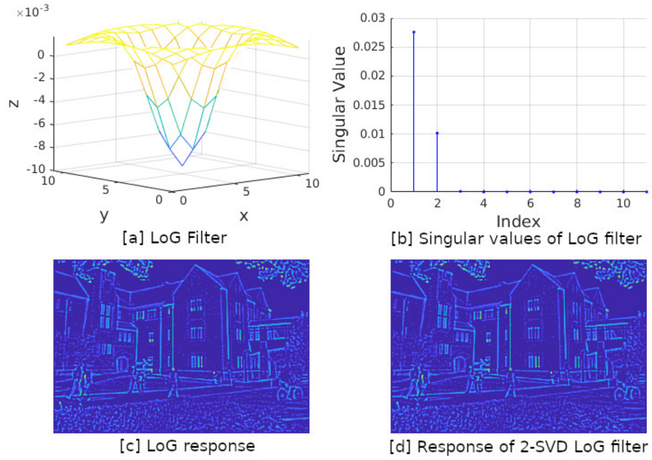

Fig. 1. [a] Visualization of the LoG filter in a 3D space. [b] Plot showing singular values decomposition of the LoG filter. It shows that the LoG can be approximated using the 2-SVD as the rest of the singular values are zeros. [c] Image showing the response of the LoG filter. [d] Image showing the response obtained from approximating the LoG filter using two separable filters of SVD.

tector [1] have long been used for detecting keypoints and image-matching tasks. Although the Harris-corner detector is invariant to certain transformations of an image, specifically, translations and rotations, it is not invariant to the scale. With the introduction of scale-space theory [11, 12] and stable keypoint detection using scale-space extrema in the Difference-of-Gaussian image [13], Lowe [2] proposed Scale-Invariant Feature Transform (SIFT) that is not only invariant to translation and rotation but also invariant to scale, noise, and illumination. Many other handcrafted keypoint detection methods like SURF[3], FAST[4], ORB [6] and BRIEF[5] have been proposed that also use a variation of separable bandpass filters for extrema detection and/or uses a different variation of a scale-space pyramid to detect the keypoints that are invariant to scale, point of view and change in lighting condition.

Our main contributions can be summarized as follows:

- We introduced the "Fast LoG" filter, a novel approach that approximates the LoG filter using two separable SVD filters for constructing the scale-space pyramid.
- We devised a method for detecting scale-space extrema by employing the polynomial fitting in the LoG response.
- Our method reduces the computational complexity by specifically computing extrema in selected pixel locations, which are determined through peak strength thresholding of LoG response.

II. Proposed Method

Our proposed method is outlined in Algorithm 1. Initially, we compute the Singular Value Decomposition (SVD) approximation of the LoG filter for a range of N different strength of σ. Values of σ_i are computed as follows:

$$\sigma_i = a\sqrt{b^i - 1} \quad (1)$$

where a and b are scaling factors and constant respectively that describes the strength of σ in the filter and $i = 1, 2, ...N$. Subsequently, we calculate the approximated LoG response for the input image to generate an LoG scale-space pyramid. Following the approach described in previous works by Lindeberg [11, 12], we scale the LoG response by σ^2. The LoG response decreases as the scale increases because the LoG operator is a high-pass filter and larger scales correspond to lower frequencies. To counteract this decrease, the LoG response is normalized by multiplying it by σ^2.

Next, we performed peak strength thresholding (P_{th}) of LoG response at different scales to eliminate the computation for the keypoints selection. We fit a third-degree polynomial to the peak responses obtained at different scales σ_i for each keypoint candidate and then calculate the first-order derivative of the curve. Subsequently, we choose the keypoint from the available candidates by ensuring that the absolute value of the first-order derivative at σ_{max} remains below the predefined gradient threshold (G_{th}).

A. LoG Scale-space Pyramid

The LoG scale-space pyramid is a multi-scale representation of an image that captures information about its structures at different scales and levels of smoothness. It is obtained by convolving the image with the LoG filter, which is the second derivative of the Gaussian function, at multiple scales. It provides a multi-scale representation of the image that can be used for keypoint detection. Keypoints are detected as the pixel location where the LoG response exceeds a certain threshold in the scale-space pyramid.

B. SVD Approximation of LoG Filter

Since the LoG filter is non-separable, directly applying the filter with a mask size of m results in a per-pixel multiplication complexity of m^2. Nevertheless, the LoG filter at scale σ can be expressed as follows:

$$LoG(x, y, \sigma) = -\frac{1}{\pi\sigma^4}\left(1 - \frac{x^2 + y^2}{2\sigma^2}\right)\exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (2)$$

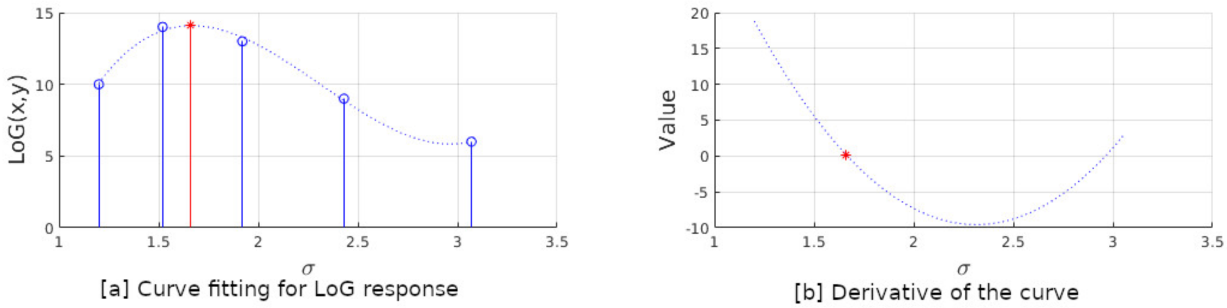

Fig. 2. [a] Curve showing the extrema detection from LoG response using polynomial fitting [b] Derivative of the curve shown in [a].

is 2-SVD separable, i.e. we can express this filter as a linear combination of two separable filters as,

$$h_0 = \sum_{i=1}^{2} \sigma_i * U_1 * V_i^T \tag{3}$$

As a result, the complexity per pixel is reduced to $4m$. To visually illustrate this concept, we provide an example in Fig. 1. Fig. 1a shows the $m \times m$ LoG filter for which we compute the SVD. In Fig. 1b, the singular values are plotted, demonstrating the effectiveness of the approach. Furthermore, in Fig. 1c and 1d, we can observe that the filter output obtained from both the original non-separable filter and the two separable filters is identical.

C. Peak Strength Thresholding

To identify keypoints, initially, we locate the extrema (minima or maxima) of the LoG response across the scale-space pyramid, leveraging the fact that keypoints typically exhibit distinct maximum or minimum LoG responses. By applying a peak strength thresholding, denoted as P_{th}, we efficiently discard pixels that are not potential candidates for keypoints. This step eliminates the need for curve fitting and significantly reduces the computational burden, particularly for large images. For pixels exceeding the P_{th} threshold, we perform a third-degree polynomial curve fitting. Next, we calculate the first-order derivative of the polynomial curve and identify sigma where the derivative tends to zero (σ_{max}), and calculate its absolute value. If the absolute value of the derivative at corresponding σ_{max} falls below the threshold G_{th}, we consider the corresponding pixel location as a valid keypoint.

Fig. 2 illustrates the curve-fitting process employed for extrema detection. In Fig. 2a, the LoG response for five different sigma values is plotted, and a third-degree polynomial curve is fitted, represented by the dotted curve. The red line corresponds to the peak response of the curve, where the gradient becomes 0 as shown in Fig. 2b, indicating a potential keypoint. To classify a response as a keypoint, we establish a threshold $G_{th} = 0.2$ and check if the derivative at σ_{max} falls within the $\pm G_{th}$ range. If it does, the response is considered a keypoint.

III. Experimental Results

A. Qualitative and Quantitative Evaluation

In order to assess the performance of our Fast LoG SIFT detector, we conducted an empirical evaluation using the pre-existing SIFT library called “VLFeat” [10] for various transformations. Fig. 3 presents a comparison between the keypoints obtained from our Fast LoG SIFT method and those generated by VLFeat SIFT. The results reveal a high correlation between the keypoints derived from Fast LoG SIFT and the keypoints evaluated in VLFeat SIFT. Fig 3a in the above comparison shows the keypoints detected using the off-the-shelf VLFeat SIFT library [10], while Fig 3b displays the keypoints detected by our proposed method.

To evaluate the robustness of our proposed method, we conducted experiments on our test image shown in Fig. 3 with different types of transformation, including images with varied resolutions, Additive White Gaussian Noise(AWGN), and gamma-correction. The objective was to evaluate the accuracy of keypoint detection in these transformed images compared to the original ones. Our findings revealed that in the case of images with AWGN of zero mean and standard deviation of 0.01, our proposed method achieved a higher accuracy of 42.42% in keypoint detection, surpassing the accuracy of the DoG-based SIFT, which was only 41.30%. Similarly, for gamma-corrected images with a Gamma value of 1.2, our method demonstrated a higher accuracy of 70.10%, outperforming the DoG-based SIFT with an accuracy of 63.33%. Additionally, when presented with images downscaled by a factor of two, our method achieved an accuracy of 46.15%, whereas the DoG-based SIFT achieved an accuracy of 38.46%. These results highlight the superior performance of our proposed method in accurately detecting keypoints in transformed images compared to the traditional DoG-based SIFT approach.

To further validate the effectiveness of our LoG SIFT detector, we utilized a synthesized image containing blobs and edges, as illustrated in Fig. 4. Upon analysis, it can be observed that the keypoints detected by our LoG SIFT method align well with the keypoints identified by the VLFeat library [10]. This provides additional confirmation of the accuracy and reliability of our proposed approach.

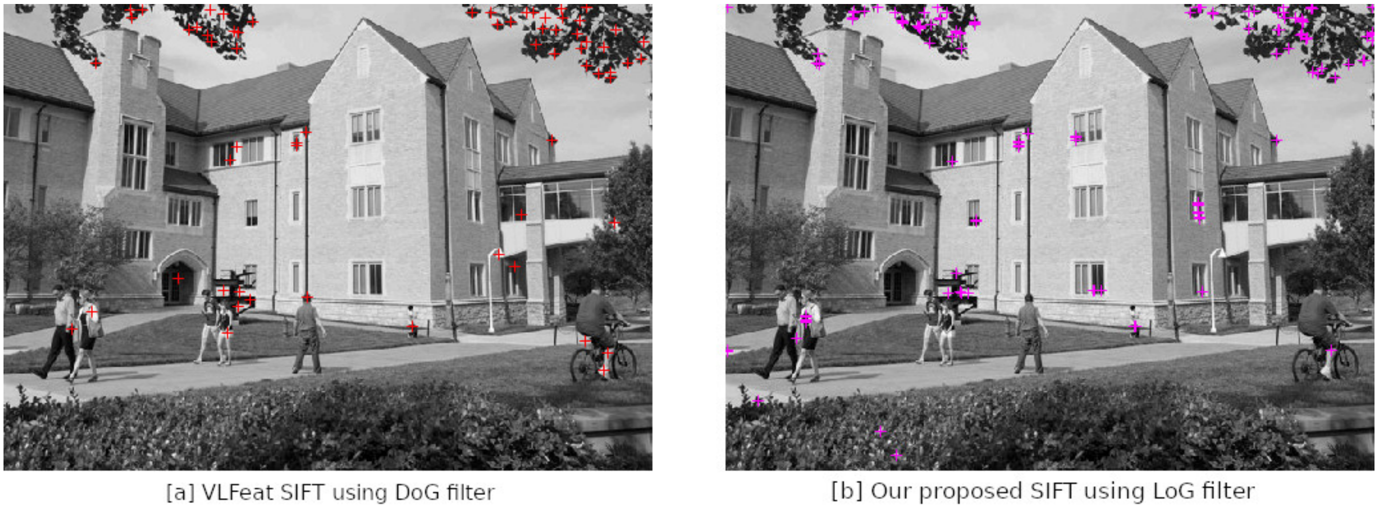

Fig. 3. Comparison of keypoint detection using SIFT using DoG filter and LoG filter. [a] SIFT keypoints detected using of-the-self library VLFeat. A total of 83 SIFT keypoints were detected.[b] SIFT keypoints detected using our proposed LoG SIFT. A total of 127 keypoints were detected for which most of the keypoints align with the VLFeat's SIFT keypoints.

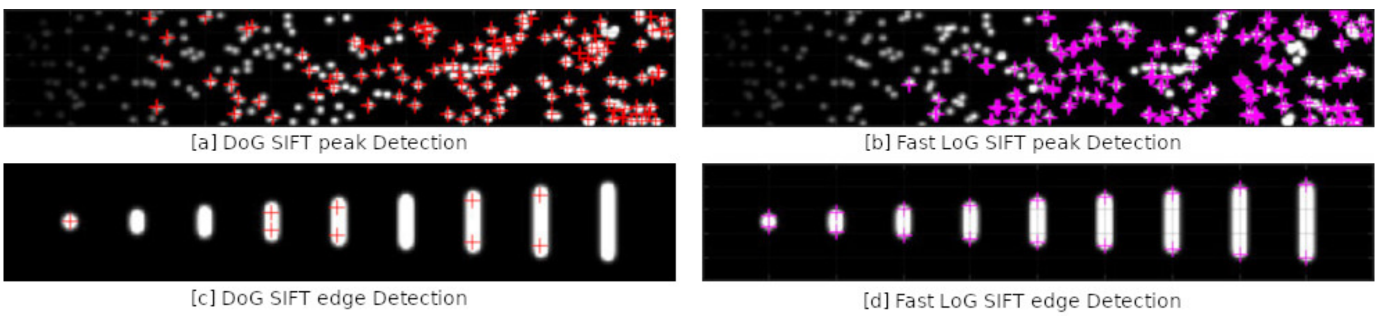

Fig. 4. Image showing detected keypoints from off-the-shelf SIFT library "VLFeat" on the synthesized image. [a] Keypoint detection in blob using DoG SIFT [b] Keypoint detection in blob using our proposed method. [c] Keypoint detection at the edge using DoG SIFT. [d] Keypoint detection at the edge using our proposed method.

B. Complexity Evaluation

Although our current implementation of the LoG SIFT detector is not fully optimized, our analysis reveals promising findings. Specifically, during the construction of the LoG feature pyramid, we have determined that it is only necessary to sample the scale-space response at four points in order to accurately fit the analytical polynomial response function. This streamlined approach translates into a multiplication complexity of $4 \times 4m$ per pixel. In comparison, the DoG SIFT method requires three octaves and five scales, each involving Gaussian filtering, resulting in a multiplication complexity of $15 \times 2m$.

Furthermore, our method incorporates LoG peak strength thresholding elimination before scale-space extrema detection, which significantly reduces the overall computation. These optimizations collectively indicate that our LoG SIFT detector will offer enhanced efficiency compared to the DoG pyramid-based solution. Through analytical assessments, we have gained confidence in the superiority of our approach in terms of speed and computational efficiency.

IV. Conclusion

In this paper, we introduce a novel approach to detect scale-invariant keypoints by employing two separable SVD filters as an approximation of the LoG filter. This separable filter significantly reduces the computational complexity, providing an order of magnitude faster implementation compared to the direct LoG computation method. Moreover, our approach retains the lower implementation overhead of the LoG pyramid-based approach compared to SIFT. The experimental results demonstrate that our method achieves a high correlation with keypoints obtained from the widely used SIFT library "VLFeat" in terms of keypoint detection. Additionally, our method offers advantages in terms of implementation complexity by avoiding unnecessary computations involved in utilizing DoG to approximate LoG and reducing the computational load during keypoint extrema detection. Moving forward, our future research will focus on improving the accuracy of keypoints by incorporating edge keypoints removal techniques. Furthermore, we will explore the assignment of descriptors to these keypoints, enhancing the overall capabilities of our proposed method.

ACKNOWLEDGMENTS

This work is supported by NSF grant 2148382 and AFRL SFFP.

REFERENCES

[1] Chris Harris, Mike Stephens, et al. A combined corner and edge detector. In *Alvey vision conference*, volume 15, pages 10–5244. Citeseer, 1988.

[2] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60:91–110, 2004.

[3] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359, 2008.

[4] Edward Rosten and Tom Drummond. Machine learning for high-speed corner detection. In *Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I 9*, pages 430–443. Springer, 2006.

[5] Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua. Brief: Binary robust independent elementary features. In *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part IV 11*, pages 778–792. Springer, 2010.

[6] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: An efficient alternative to sift or surf. In *2011 International Conference on Computer Vision*, pages 2564–2571, 2011. doi: 10.1109/ICCV.2011.6126544.

[7] Edward Rosten and Tom Drummond. Machine learning for high-speed corner detection. In *Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I 9*, pages 430–443. Springer, 2006.

[8] Stefan Leutenegger, Margarita Chli, and Roland Y Siegwart. Brisk: Binary robust invariant scalable keypoints. In *2011 International conference on computer vision*, pages 2548–2555. Ieee, 2011.

[9] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 886–893. Ieee, 2005.

[10] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms. `http://www.vlfeat.org/`, 2008.

[11] Tony Lindeberg. Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention. *International Journal of Computer Vision*, 11(3):283–318, 1993.

[12] Tony Lindeberg. Scale-space theory: A basic tool for analyzing structures at different scales. *Journal of applied statistics*, 21(1-2):225–270, 1994.

[13] David G Lowe. Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on computer vision*, volume 2, pages 1150–1157. Ieee, 1999.