An Efficient Technique to Extricate Keypoints from a digital image

Rachel Selva Dhanaraj, and M Sridevi
Department of Computer Science and Engineering, National Institute of Technology, Tiruchirapalli, Tamil Nadu, India.

Abstract: With the immense technology emerging in this period, the transmitted image information across the network can be confined easily, maneuver, and then disseminated. For any labor done on processing images, it is mandatory to detect suitable meaningful keypoints. There are existing methods that perceive key points from an image but they fail to detect key points from a smooth textured area or they do not detect enough keypoints. A method is proposed using an improved version of the Scale Invariant Feature Transform (SIFT) technique that is capable of diagnosing key points even in smooth areas as well as administer key points towards the whole image evenly.

Keywords: Keypoint, Extraction, Scale Invariant Feature Transform, Scale Space, Gaussian operator.

1. INTRODUCTION

In this era, digital images play a vital role in our habitual life. Digital images are transmitted across the globe using profuse mediums of transmission. These images are thwarted and tampered by intruders before attaining the destination. Image manipulation tools are readily available that can distort the very uniqueness of digital images. Images are comprehensively discharged across social media which are reoriented easily to misrepresent their indigenous meaning with malevolent intention. This is the pre-vision of major consequences, as in the instance of the present scenario where everything is digitized especially in the defense sector, judiciary, media, astronomy, etc. It is the need of the hour to identify such forgeries that have occurred in the digital images. It is an immense confrontation to validate such images that have undergone amendments through distortion.

In the present generation, multitudinous technologies were at the forefront to identify the forged portion of an image. The real matter underlines in pinpointing the appropriate and meaningful key points in the image to discern and to get the authentic feature descriptors. These feature descriptors would succour to detect forgery, to recognize objects, to identify outliers, etc. from an image.

The objective of this work is to distillate the meaningful key points from an image in an efficient manner. The existing mechanisms like Scale Invariant Feature Transform (SIFT) [10], Speeded Up Robust Features (SURF) [11], Binary robust invariant scalable keypoints (BRISK) [12] have been enabled to detect key points. Unfortunately, these methods are unable to extract all consequential key points. The proposed method applies permutated SIFT version of keypoint extraction.
1.1. Organization
This paper is organized as follows. Section 2 provides an overview of the related research work. Section 3 presents the Proposed System Performance Evaluation and Results is discussed in section 4. This paper is concluded in section 5

2. RELATED WORK

SURF technique can competency to perform a fast and robust algorithm. It proffers the use of Gaussian scale-space by approximating Gaussian derivatives through box filters [1]. It employs basic Hessian matrix approximation due to its better execution. This approach is devoid of preserving the object boundaries and fails to associate all interest points that are a prerequisite of bargaining the desired output.

SIFT is attended in [10] that fosters a scale space and enumerate the Difference of Gaussians (DoG). Several octaves of the initial image are generated. The image size of each octave is halved the preceding one. Gaussian Blur operator is sought to blur each image progressively within an octave. The difference between two consecutive scales furnishes a DoG. Approximate maxima and minima are distinguished in DoG by collating the neighboring pixels that constitute the valid key points. These key points are approximate as maxima and minima almost do not lie exactly on the pixel.

In [12], the authors have proposed a novel method to effectuate interest points. The saliency criterion is used to detect mandatory key points from both the image and scale dimensions. From the image pyramid, interest points are deduced in octave layers furthermore in layers in-between to procure and enhance the effectiveness of computation. Each keypoint location and scale is ascertained in the continuous domain by dint of quadratic function fitting.

The authors in [13] have dispensed a machine learning technique Features from Accelerated Segment Test (FAST) for dredging key points. It is an extravagant, high-standard corner detector. It utilizes accelerated segment test criteria to create a corner detector that would detect corners from a set of images. This method is not robust to noise and is reliant on the threshold. [14] Contemplates a significant technique that escalates the effectiveness of FAST. It is achieved by generating an ideal decision tree in enlarged configuration space. An adaptive and generic accelerated segment test is thus yielded by amalgamating various specialized trees.

The authors in [3] used binary strings which are computed from image patches as an effective interest point descriptor called BRIEF. The intensities of pairs of points are contrasted to obtain individual bits. Comparison is tendered by computing Hamming distance. ORB (Oriented Fast and Rotated BRIEF) [2] is a brisk keypoint detector inspired by BRIEF. ORB adds a fastened and accurate orientation component and computes BRIEF features methodically. ORB is consequently rotation invariant and immune to noise.

The authors in [4] instigate KAZE features that employ nonlinear diffusion filtering to relate and portray 2D features in nonlinear scale space. Along these lines, the acclimatization of blurring provincially is applied to the image data that reduces noise but sustains object boundaries. In [5], a mathematical structure namely Fast Explicit Diffusion (FED) is implemented to recognize features at a drastic rate in nonlinear scale space. To prompt keypoint detection, the paper kindles to enclose the FED scheme in a pyramidal framework with a fine to coarse approach. These features are entitled as Accelerated-KAZE. Hence, this technique procures low computationally demanding key points that endorse the merits of nonlinear diffusion filtering.

Neoteric approach is proposed in [1] that discerns key point by traversing the information within the descriptor space. Describe-to-Detect (D2D) describes the image and then later perceives keypoints. The selection of key points is accomplished by choosing high information content from the
descriptors defined. It is done by computing the entropy of the descriptor or either by autocorrelation. The location that has high entropy or correlation would then be reckoned as a key point.

The paper [6] inaugurates an algorithm that utilizes combinations of various key point extraction techniques to reduce a substantial amount of key points. This method draws 2 step filtering process: spatial filtering to minimize interest points by employing entropy and adopting significant points with is maximal information content.

In [7], the author pioneers an unsupervised deep learning-based keypoint detection and descriptor. Key point scores and positions are cultivated automatically by exploiting the Siamese network and novel loss function. The outcome of keypoint detection and descriptor is Unsuperpoint. Regression is used to form Unsuperpoint end-to-end trainable and to integrate non-maximum suppression in the model.

3. PROPOSED METHODOLOGY

The proposed keypoint extraction purpose utilized by numerous block-based or feature-based algorithms likewise fiercely exposes several key points or none at smooth textured locations in an image. The proposed method envisages overwhelming the issue of generating valid interest points. The technique pivots on the SIFT perspective that is surpassed to detect and distinguish the interest points.

SIFT attributes can be confronted at various spaces employing scale-space representation executed as an image pyramid. The levels of the pyramid are retrieved by Gaussian smoothing and subsampling of the image resolution. Keypoints are validated by taking into consideration of local maxima and minima in the scale space.

3.1. Detection of Extrema in Scale Space

SIFT instigates blurred out images progressively from the commencing image. The rescaling of the size of images is perpetrated and blurring is executed auxiliary. The convolution of the Gaussian operator on the image is called blurring.

A 2D image represented is defined as given in eq. (1)

$$L(x,y,\sigma) = G(x,y,\sigma) \ast I(x,y)$$

(1)

where G is the Gaussian Blur operator and “∗” is the convolution operation. The Gaussian Blur operator is given in eq. (2)

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/(2\sigma^2)}$$

(2)

where σ is the scale parameter and (x,y) is a spatial coordinate.

The difference of Gaussians (DoG) is precipitated from the blurred images. The outcome in the form of images from DoG is of great convenience for enabling to distinguish interest points in the image. Endorsement of maxima and minima in the DoG images is attained by juxtaposing the neighboring pixels with all the three scales i.e current, above, and below. These local extrema points are scrutinized as a key point.

Gradient directions and magnitudes are collected around each key point that is calculated by using eq. (3) and (4) respectively

$$m(x,y) = \left( (L(x + 1, y) - L(x - 1, y))^2 + (L(x, y + 1) - L(x, y - 1))^2 \right)^{1/2}$$

(3)
\[ \theta(x,y) = \tan^{-1}\left(\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)}\right) \] (4)

3.2 Normalized and Robust Threshold Selection Approach

SIFT detector removes low dissimilitude interest points from the image by recruiting threshold. However, these low contrast keypoints cannot be avoided as there is a possibility of information being embedded in it. Hence, a normalized and robust threshold selection approach [8] is approved that would dodge the consequence of perpetuating low contrast keypoints.

The novel image is considered and subdivided into \( n \times n \) sub-images. With the aid of the SIFT algorithm, authentic keypoints are generated with an appropriate threshold worth. Then, the keypoints scattering level called Keypoint Uniformity Measurement (KUM) is defined and premeditated as shown in the algorithm. The threshold \( \lambda \) that is purged of the low contrast point is customized by KUM [8].

If KUM value \( \phi \) is less than \( \tau \), the keypoints are equivalently dispensed otherwise the threshold value is refurbished as shown in the algorithm. A variable \( t \) is used to restrict the number of iterations which is synchronized to 4. \( n \) is initialized to 10 and \( \tau \) is set to 0.3.

3.3 Algorithm 1 – Generation of Keypoints using Keypoint Uniformity Measurement

1. Input the original image \( I(x,y) \)
2. Initialize \( t \) as 1
3. Partition the input image into \( n \times n \) sub-image
4. Spot the keypoints by applying the SIFT algorithm with threshold \( \lambda \)
5. Determine the keypoint uniformity measurement (KUM) value \( \phi \) as follows:
   5.1. Describe a matrix \( k(i,j) \) and \( k_{i,j} \) is the keypoints number of sub-image
   5.2. Initialize \( T \) as the number of identified keypoints in the image \( I \).
   5.3. Standard keypoints number \( s \) is computed using
   \[ s = T\left(\frac{M}{n}\right) \times \left(\frac{N}{n}\right)^{-1} \] (5)
   5.4. KUM is computed as follows
   \[ \phi = \left(\frac{\sum_{i=1}^{[M/n]} \sum_{j=1}^{[N/n]} (k_{i,j} - s)^2 \times T^{-1}}{T^{-1}}\right)^{1/2} \times T^{-1} \] (6)
6. If (KUM < \( \tau \) or \( t > 4 \)) then
   • Output the identified keypoints
   • Else
     • \( \lambda = \lambda/3 + \text{mod}(\lambda,3) \)
     • \( t = t+1 \)
     • return to step 4

4. RESULTS AND PERFORMANCE EVALUATION

The proposed algorithm is implemented using Matlab and the results are presented in this section. CoMoFoD database [9] is used for the experimenting the proposed algorithm. It consists of 260 image sets, 200 images in the small image category (512 x 512), and 60 images in the large image category (3000 x 2000).

For implementation, the proposed method is compared with state-of-the-art methods like SIFT. The proposed method detects more keypoints at uniform and textureless area and distributed evenly than the state-of-the-art method. It is evident from the results shown in Fig. 1 to Fig. 4.
The performance is evaluated using threshold selection in the proposed system and SIFT techniques and is depicted in Table 1. The proposed system tends to give more number of valid keypoints than other state-of-the-art method.
Table 1: Comparison of Performance of the Proposed system with SIFT

| Threshold (λ) | Proposed System | SIFT  |
|--------------|-----------------|-------|
| 0.1          | 3146            | 1126  |
| 0.085        | 3443            | 1371  |
| 0.075        | 3633            | 1575  |
| 0.05         | 4118            | 2110  |
| 0.025        | 4741            | 2838  |

5. CONCLUSION

The perception of the key points from an image could be an exigent task on the condition that the image is smoothened or textureless. Through the inauguration of a novel algorithm called Normalized and robust threshold selection approach, it overcomes the challenge of diagnosing the key points. This modus operandi utilizes key point uniformity measurement to circumvent low contrast key points to be eradicated from the image. The anatomization results conclude that this methodology is better than the state-of-the-art techniques such as SIFT and Harris detector through precise detection of keypoints from smoothened or uniform areas of the image. The proposed method is more overwhelming to detect the inadequacy of key points from the desired image. In future, we would work to generate descriptors for the keypoints that would be invariant to rotation, scaling, viewpoint and mirror reflection.

REFERENCES

[1] Yurun Tian, Vassileios Balntas, Tony Ng, Axel Barroso-Laguna, Yiannis Demiris, Krystian Mikolajczyk, “D2D: Keypoint Extraction with Describe to Detect Approach”, Computer Science, Engineering, ArXiv 2020.
[2] Ethan Rublee, Vincent Rabaud, Kurt Konolige, Gary Bradski, “ORB: An efficient alternative to SIFT or SURF”, IEEE International Conference on Computer Vision (ICCV 2011).
[3] Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua, “BRIEF: Binary Robust Independent Elementary Features”, Computer Vision – ECCV 2010.
[4] Alcantarilla P.F., Bartoli A., Davison A.J. (2012) KAZE Features. In: Fitzgibbon A., Lazebnik S., Perona P., Sato Y., Schmid C. (eds) Computer Vision – ECCV 2012. ECCV 2012. Lecture Notes in Computer Science, vol 7577. Springer, Berlin, Heidelberg.
[5] Pablo Alcantarilla, Jesus Nuevo, and Adrien Bartoli, “Fast explicit diffusion for accelerated features in nonlinear scale spaces”, British Machine Vision Conference (BMVC) 2013, DOI: 10.5244/C.27.13.
[6] Maja Rudinac, Boris Lenseigne, Pieter Jonker, “Keypoint extraction and selection for object recognition”, MVA2009 IAPR Conference on Machine Vision Applications, May 20-22, 2009, Yokohama, JAPAN.
[7] Peter Hviid Christiansen, Mikkel Fly Kragh, Yury Brodskiy, Henrik Karstoft, “Unsuperpoint: End-to-End unsupervised interest point detector and descriptor”, arXiv:1907.04011v1 [cs.CV] 9 Jul 2019.
[8] Yang, B., Sun, X., Guo, H. et al. A copy-move forgery detection method based on CMFD-SIFT. Multimed Tools Appl77, 837–855 (2018).
[9] Tralic D., Zupancic I., Grgic S., Grgic M., "CoMoFoD - New Database for Copy-Move Forgery Detection", in Proc. 55th International Symposium ELMAR-2013, pp. 49-54, September 2013.
[10] David G Lowe. Distinctive image features from Scale-Invariant keypoints, 2004
[11] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-Up robust features (SURF).Comput. Vis. Image Underst., 110(3):346–359, June 2008.
[12] Stefan Leutenegger, Margarita Chli, and Roland Siegwart. BRISK: Binary robust invariant scalable keypoints. In 2011 IEEE international conference on computer vision (ICCV), pages
2548–2555, 2011.

[13] Rosten E., Drummond T. (2006) Machine Learning for High-Speed Corner Detection. In: Leonardis A., Bischof H., Pinz A. (eds) Computer Vision – ECCV 2006. ECCV 2006. Lecture Notes in Computer Science, vol 3951. Springer, Berlin, Heidelberg.

[14] Mair E., Hager G.D., Burschka D., Suppa M., Hirzinger G. (2010) Adaptive and Generic Corner Detection Based on the Accelerated Segment Test. In: Daniilidis K., Maragos P., Paragios N. (eds) Computer Vision – ECCV 2010. ECCV 2010. Lecture Notes in Computer Science, vol 6312. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-15552-9_14.

[15] Yurun Tian, Bin Fan, and Fuchao Wu. L2-net: Deep learning of discriminative patch descriptor in euclidean space. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 661–669, 2017.

[16] Nikolay Savinov, Akihitoseki, Lubor Ladicky, Torsten Sattler, and Marc Pollefeys. Quad-networks: Unsupervised learning to rank for interest point detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1822–1830, 2017.

[17] K.Mikolajczyk, T.Tuytelaars, C.Schmid, A.Zisserman, J.Matas, F.Schaffalitzky, T.Kadir, 1L.Van Gool: "A Comparison of Affine Region Detectors", International Journal of Computer Vision 65(1-2), 43-72, 2005.

[18] M. Stark and B. Schiele, "How good are local features for classes of geometric objects", International Conference on Computer Vision, 1-8, 2007

[19] S. Salti, A. Lanza and L. Di Stefano, "Keypoints from symmetries by wave propagation," 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, 2013, pp. 2898-2905, DOI: 10.1109/CVPR.2013.373

[20] K. M. Yi; E. TrullsFortuny; V. Lepetit; P. Fua, "LIFT: Learned Invariant Feature Transform", European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, October 8-16, 2016. p. 467-483. DOI : 10.1007/978-3-319-46466-4_28.

[21] J. Aldana-Iuit, D. Mishkin, O. Chum, and J. Matas. In the saddle: chasing fast and repeatable features. In 2016 23rd International Conference on Pattern Recognition (ICPR), pages 675–680. IEEE, 2016.

[22] V. Balntas, K. Lenc, A. Vedaldi, and K. Mikolajczyk. Hpatches: A benchmark and evaluation of handcrafted and learned local descriptors. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), volume 4, page 6, 2017.

[23] V. Balntas, E. Riba, D. Ponsa, and K. Mikolajczyk. Learning local feature descriptors with triplets and shallow convolutional neural networks. In British Machine Vision Conference (BMVC), volume 1, page 3, 2016.

[24] A. Benbihi, M. Geist, and C. Pradalier. Elf: Embedded localisation of features in pre-trained cnn. In Proceedings of the IEEE International Conference on Computer Vision, pages 7940–7949, 2019.

[25] J. Revaud, P. Weinzaepfel, C. De Souza, N. Pion, G. Csurka, Y. Cabon, and M. Humenberger. R2d2: Repeateable and reliable detector and descriptor. arXiv preprint arXiv:1906.06195, 2019.

[26] L. Zhang and S. Rusinkiewicz. Learning to detect features in texture images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 6325–6333, 2018.

[27] H. Noh, A. Araujo, J. Sim, T. Weyand, and B. Han. Large-scale image retrieval with attentive deep local features. In Proceedings of the IEEE International Conference on Computer Vision, pages 3456–3465, 2017.