2022 The 22st International Conference on Control, Automation and Systems (ICCAS 2022)
BEXCO, Busan, Korea, Nov. 27∼Dec. 01, 2022

EventPointNet: Robust Keypoint Detection with Neuromorphic Camera Data

Howoong Jun[1], Sangil Lee[2], and Songhwai Oh*[3]

[1]Graduate School of Artificial Intelligence (GSAI), Seoul National University,
[2]Department of Mechanical and Aerospace Engineering (MAE), Seoul National University,
[3]Department of Electrical and Computer Engineering (ECE) and Graduate School of Artifical Intelligence (GSAI), Seoul National University,
Seoul, 08826, Korea (howoong.jun@rllab.snu.ac.kr, sangil07@snu.ac.kr, and songhwai@snu.ac.kr)* Corresponding author

Abstract: In this paper, we propose a new method for keypoint detection using neuromorphic camera data. Robust keypoint detection in diverse conditions is a major issue for visual simultaneous localization and mapping (SLAM), place recognition, and computer vision. Recently, many methods adopt supervised learning to solve the problem. However, it is hard to define the exact reference keypoints on natural scenes, so the training process can be ambiguous for the problem. To handle this issue, we propose a new method named EventPointNet which is trained from data collected from a neuromorphic camera, also known as an event-based camera. Since the event-based camera captures natural edge points from any scenes regardless of illumination and viewpoint changes, the data can be used as proper references for keypoint detection. Therefore, a network trained with these data can detect distinct keypoints on a gray-scale image captured from a conventional camera. The proposed method is validated by comparing with both handcrafted and learning-based approaches on HPatches dataset. The experimental results show that EventPointNet detects more valid keypoints than the other methods in terms of both qualitative and quantitative results, especially on the illumination conditions with 1.31% higher matching score compared to the second best method. We also perform the visual odometry experiments on the KITTI dataset to show that EventPointNet can be applied to robotic applications. In particular, EventPointNet shows a reduction of 30.74% in the trajectory error compared to the second best algorithm.

Keywords: Keypoint, Local Feature, Visual Localization

1. INTRODUCTION

Keypoint detection is one of the main problems across several robotics fields such as visual simultaneous localization and mapping (SLAM) [1]. In particular, it is the starting point for a robot to understand its surrounding environment. The definition of the problem is to find distinct local features that can be detected repeatably from different images of the same scene. In other words, the same point should be detected from images with different viewpoints, illumination, and camera models. Even though the effect of different camera models is minimal due to the significant improvement of hardware technology, illumination and viewpoints still remain as the main problem.

Many researchers try to solve the problem with both learning-based and handcrafted approaches. Mainly, supervised learning methods have been actively investigated since the convolutional neural network has made substantial performance improvement on image-related problems in the last decade. These methods utilize synthetic image data or self-supervised approach with ground truth correspondence to figure out the keypoint detection problem. However, determining the reference keypoints on natural scenes for training data can be obscure since there is no explicit definition to it. Furthermore, it is hard to train images with low contrast or low illumination since there are no precise reference keypoint data. Therefore, new reference data that can cover diverse illumination and viewpoint conditions of natural scenes are needed for training a robust keypoint detector.

In this paper, we propose a new keypoint detector using neuromorphic camera data. A neuromorphic camera, also known as an event-based camera, detects the illumination change of individual pixels. The sensor outputs an event consisting of a timestamp, pixel position, and polarity value. A pixel position is the 2D coordinate of a pixel in an image whose brightness has changed. A polarity value is an 1-bit indicator that describes whether the pixel is getting brighter (ON) or darker (OFF) [2]. By means of this nature, the output of the sensor includes natural edge points of the scene extracted along with the camera and object movements. Also, due to its high dynamic range, it can collect visual data even in a scene with low illumination. This motivated us to use the event data as references for training a keypoint detector.

The proposed method, which is named EventPointNet, is a supervised-learning-based keypoint detector trained with event data. In order for EventPointNet to achieve the desired result, we modify the raw event data into candidate keypoints. Due to the characteristic of the event camera, reference feature data are repeatably collected from images with both low and high illumination. Since data are collected with natural images, the method can easily adapt to real-world images without any additional adaptation method.

We conduct three experiments, keypoint detection, keypoint matching, and visual odometry, to show that EventPointNet can be used as an alternative for keypoint-based applications. The keypoint detection and keypoint matching experiments show that the proposed method performs better than existing methods on the HPatches

dataset. On the visual odometry experiment, EventPointNet shows the best results in terms of the absolute trajectory error on the KITTI dataset [3], showing a reduction of 30.74% in error compared to the second best algorithm.

2. RELATED WORK

2.1 Handcrafted Methods

Diverse approaches to solve keypoint detection problems have been studied for many decades. Traditional approaches utilize local geometric information to find invariant points. One of the most well-known traditional local features is scale-invariant feature transform (SIFT) [4]. SIFT is a patch-based keypoint detecting method that uses the difference of the Gaussian blur image and original image to find keypoint location independent of scale. It accumulates the differences and finds the pixel that shows the most significant difference. Despite its various uses, it has the disadvantage of consuming high costs. On the other hand, oriented FAST and rotated BRIEF known as ORB [5] tries to reduce the computation for keypoint detection problems. It combines FAST [6] keypoint detector and BRIEF [7] descriptor to achieve good performance with low cost to overcome the limitations of SIFT. BRISK [8] creates scale-space for keypoint detection so that it can generate scale-invariant robust feature points. Also, to achieve the rotation-invariant local feature description, BRISK identifies the direction of each detected keypoints. In the case of KAZE [9], it uses a non-linear diffusion filter instead of a Gaussian filter for keypoint detection. Compared with the Gaussian filter-based methods, KAZE conserves details or edges of the image data so that it can find more robust keypoints. Accelerated-KAZE, known as AKAZE [10], tries to reduce the computation of KAZE. It replaces additive operator splitting into fast explicit diffusion to boost up the speed.

2.2 Learning-Based Methods

Recently, almost every method focuses on using deep learning due to its high performance on the image. Superpoint [11] is a self-supervised keypoint detecting and description method. Firstly, they use a pre-trained keypoint detector called MagicPoint which uses synthetic images for training the network. This MagicPoint and homographic adaptation method make it possible to perform robustly in detecting and describing keypoints. R2D2 [12] focuses on reliability of the keypoints. Its network has an independent branch of reliability network so that it can obtain the reliability map. In this manner, instead of the detect-then-describe paradigm, they use the detect-and-describe approach that outputs detectors and descriptors simultaneously. LF-Net [13] is another local feature detection method with supervised learning. It requires a camera pose, calibration parameters, and depth map for training local features. Also, its descriptor is based on the image patches around the chosen keypoints. Key.net [14] combines handcrafted and learned features and proposes light and efficient keypoint detector. Instead of using supervised learning to the raw input image, the method encodes it with handcrafted methods and then applies CNN-filters.

3. EVENTPOINTNET

The main purpose of EventPointNet is to find keypoints that can be repeatably detected and uniquely described from images under various conditions. For this purpose, we use neuromorphic camera data known as event data as reference keypoints. Event camera captures the natural edge points of the scene. In addition, due to the characteristics of the event camera, the scene can be detected repeatably regardless of light condition. This proves that the event data can be valuable reference data for training both dark and bright images.

3.1 Event Data

We use event data captured from neuromorphic camera for training the network. Since each pixel value of the event data indicates how many times it is exposed for a certain period of time, we use locations of the pixels that contain event data as possible keypoints. Also, we modify the values of the event pixels to use them as candidate keypoints. The polarity of the event data is not considered since both positive and negative data are all necessary for the process. The event data is collected for Δt, which is a time interval between two pixel images, and normalized based on the exposure of the event data. Therefore, each event pixel can be regarded as a probability value for a possible keypoint. For example, for a pixel that is the most frequently sensed by a dynamic vision sensor (DVS) during Δt, the event value is set to one.

With the data prepared for training EventPointNet, the training image is encoded with the CNN-filter The network outputs heatmap of the possible keypoints and then loss is computed between the heatmap and the candidate keypoints extracted from the event data. Note that the event data are only used for training the network. With the trained network, EventPointNet can infer keypoint result of a gray-scale image captured from a conventional camera.

3.2 Network Architecture

The overall network structure for the proposed keypoint detector is described in Figure 1. The backbone structure is mostly ascribed from [11], which is a VGG-like architecture [15]. The architecture includes 3×3 convolution layers and 2×2 sized max pooling layer on every two convolution layers. Therefore, three max-pooling layers reduce the image size into $1/64$. Last convolutional layer modify the output of the encoder into 65 channels. A channel-wise softmax is applied to the output, and then it is up-scaled into the original image size. One channel which indicates 'no keypoint' is removed during the up-scaling process. The other 64 chan-

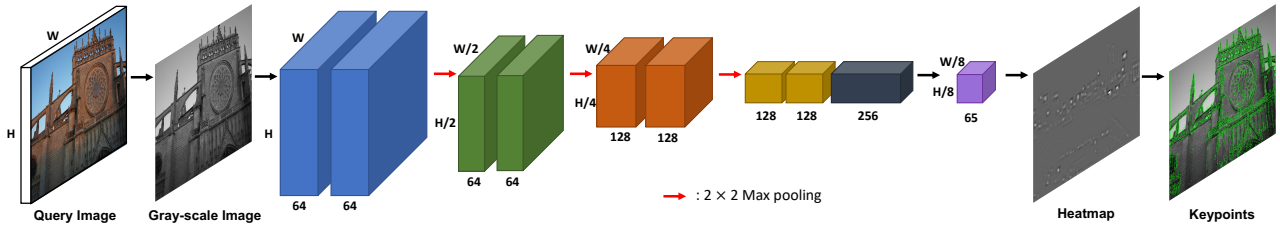

Fig. 1. An inference process of the EventPointNet keypoint detector. The network encodes the input image captured from a conventional camera into $W/8 \times H/8 \times 65$ and derives a heatmap for possible keypoint locations. With the information in the heatmap, keypoints can be generated.

nels are up-scaled into 8×8 pixel. Finally, the output of the overall network is a heatmap of possible keypoints. All convolution layers are followed by a rectified linear unit (ReLU) activation function. For the keypoint detector loss function, mean-squared error (MSE) loss is used. We apply non-maximum suppression (NMS) for the final output.

4. EXPERIMENTS

In this section, we present implementation details and experimental results. We conduct three experiments, keypoint detection, keypoint matching, and visual odometry. The overall experiments are compared with both handcrafted and learning-based methods. For the handcrafted methods, AKAZE [10], BRISK [8], KAZE [9], ORB [5], and SIFT [4] are used and for the learning-based methods, Superpoint [11], R2D2 [12], and LF-Net [13] are used for the experiments. Handcrafted methods are implemented with the pre-defined functions in the OpenCV framework. For learning-based methods, we use the pre-trained models and codes provided by the authors. Since R2D2 has three pre-trained models, we use the model named `r2d2_WAF_N16.pt`.

4.1 Implementation Details

The overall environment for the experiments can be pulled at docker hub repository[1]. Also, the code and trained parameters for the EventPointNet are available at Github[2].

4.2 Training Data

We use a multi-vehicle stereo event camera (MVSEC) dataset [16] for training the network. MVSEC dataset contains event data and gray-scale images, which are collected with DAVIS346B event camera. Image sizes of both event and gray-scale image data are 346×260. Since DAVIS346B collects both event and gray-scale intensity data at the same pixel simultaneously, it does not need timestamp synchronization or extrinsic camera calibration between them.

However, there is a problem with using MVSEC dataset directly for training EventPointNet. Since the DVS cannot sense anything if there is no difference between two frames, the event image returns no information on motionless scenes. For instance, if the camera and objects in the scene do not move, all event values in the image are set to zero. Unfortunately, MVSEC dataset contains these stationary scenes such as traffic jams, traffic lights, and parking, which can interfere during training. Therefore, we manually filter out stationary frames from the training set.

Additionally, since data are collected on a vehicle, which mainly contains horizontal movements, the event camera mostly captures the vertical edges of the scene. To avoid this data bias, we augment data with random rotational angles. Also, to handle various light conditions, we vary the brightness of the gray-scale image and include them in the training data.

4.3 Keypoint Detection

To evaluate the proposed keypoint detection method, we use the HPatches dataset [17], which is widely used for evaluating keypoint-detection-related problems. For a fair comparison, we resize the input images of HPatches dataset into 320×240 and use a maximum of 1,000 keypoints per image based on the scoring or confidence value provided by the method to maintain the same density of keypoints on various images. In case of methods that do not provide scoring or confidence value, we randomly pick 1,000 points to compute the results.

The qualitative result for the experiment is shown in Figure 2. The image used in Figure 2 is an example of the HPatches dataset, which contains both bright and dark sides in the same image. EventPointNet tends to derive rich keypoints regardless of illumination changes compared with the other methods, especially handcrafted methods. Interestingly, only EventPointNet detects keypoints on extremely dark sides of the image which are located on the bottom and top-right side of the image. Detecting keypoints on dark sides can be beneficial on diverse applications such as visual localization and visual odometry since they need to find keypoints on images with low illumination such as images taken at night.

Table 1 shows the repeatability results for keypoint detection. Repeatability (r) indicates the ratio of the number of corresponding keypoints (n_c) to the minimum number of keypoints between two images ($\min(n_1, n_2)$) [18]. The result shows that the proposed method achieves the best repeatability on the viewpoint and slightly lower repeatability on the illumination. Interestingly, handcrafted methods such as BRISK and ORB perform well

[1] rllabsnu/eventpointnet:base
[2] https://github.com/rllab-snu/eventpointnet.git

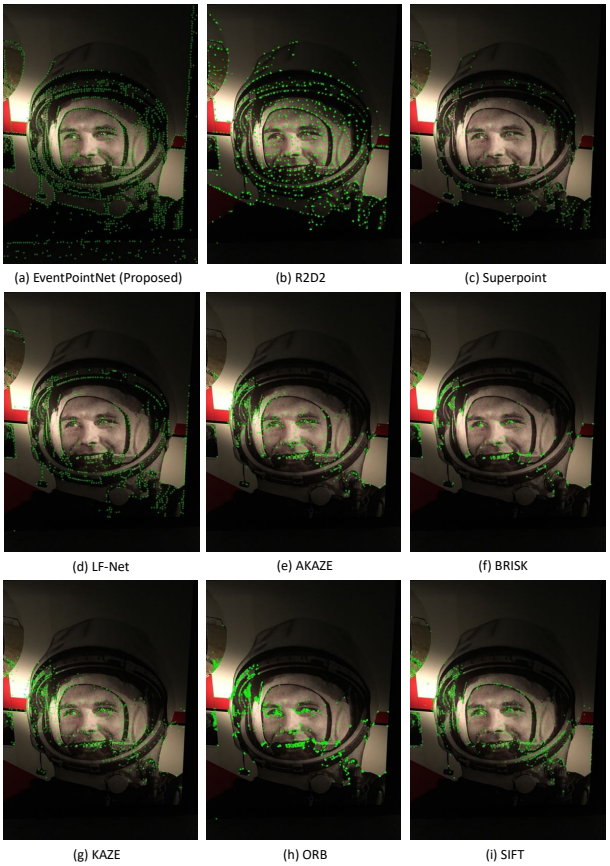

Fig. 2. Keypoint detection results with `v_yuri` data of HPatches dataset.

on the illumination dataset. The reason for this is because BRISK and ORB tend to derive patchy and clumped keypoints, so if there exists a correctly matched region, a large number of overlapping corresponding points will be detected inside it. This characteristic can provide a positive result in the repeatability experiment, but in terms of camera pose estimation, it is not functional, which is shown in Section 4.4 and Section 4.5. Aside from BRISK and ORB, EventPointNet outperforms other methods on illumination condition by reducing 28.41% error rate compared to the second best method, AKAZE. In addition, as illustrated in Figure 2, R2D2 also detects keypoints on dark sides of the image as much as EventPointNet. However, the quantitative result shows that the proposed method finds more repetitive keypoints than R2D2 on both viewpoint and illumination changes.

4.4 Keypoint Matching

We conduct further experiments about feature matching with the HPatches dataset and the methods used in the keypoint detection experiment. The SIFT description method is used as a descriptor for the proposed method. Lastly, Brute-force algorithm [19] is used for feature matching.

The qualitative result for the feature matching experiment is shown in Figure 3. To show that the proposed method can detect robust keypoints on the low illumination images, we select data with illumination changes of

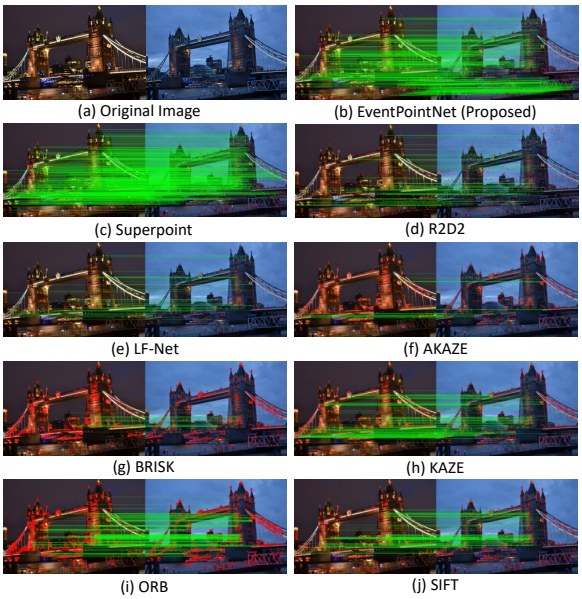

Fig. 3. Keypoint matching results with `i_londonbridge` data of HPatches dataset.

Table 1. Repeatability result and matching score (%)

	Repeatability		Matching Score	
	View	Illum	View	Illum
EventPointNet	**44.88**	83.19	48.44	**83.65**
Superpoint [11]	26.40	74.49	**56.82**	82.57
R2D2 [12]	29.88	76.02	49.39	79.59
LF-Net [13]	40.87	76.39	40.55	70.01
AKAZE [10]	31.24	76.52	57.73	72.11
BRISK [8]	42.26	**85.52**	41.43	53.21
KAZE [9]	32.95	75.36	56.62	67.23
ORB [5]	35.56	83.71	46.21	54.17
SIFT [4]	40.59	68.54	51.90	57.09

the natural scene as an example (`i_londonbridge`). The result shows that the proposed method detects valid keypoints distributed on the surface of the relevant objects and shows outstanding matching results. We apply random sample consensus (RANSAC) on every methods to filter out outliers for matching images.

For quantitative analysis, we employ the concept of matching score used in [20], which is the ratio of correct matches and correspondences. As we know the reference matches from the homography matrix given by the dataset, we can evaluate whether the estimated matches are correct or not. If the distance between the estimated matching point and reference matching point is below the threshold pixel distance, the two points are regarded as a correct match. We apply the same resize resolution and maximum threshold parameters of the repeatability test for all tested algorithms. Since input images are resized, we apply scale matrix for the given homography matrix.

To evaluate the performance of the raw key point match results, we do not apply RANSAC in the quantitative experiment. Since the proposed method mainly focuses on the keypoint detection part, we apply the same descriptor (SIFT descriptor) to all methods for a fair com-

Table 2. Relative Pose Error (RPE) and Absolute Trajectory Error (ATE) of KITTI Dataset

Data	Error Metric	**EPN (Proposed)**	Superpoint [11]	R2D2 [12]
00	RPE	1.926	1.917	1.974
	ATE	41.850	104.010	142.169
01	RPE	4.592	4.370	4.318
	ATE	386.251	847.681	431.631
02	RPE	2.399	2.409	2.366
	ATE	100.729	132.901	326.387
03	RPE	1.775	1.776	1.771
	ATE	24.477	38.529	37.751
04	RPE	3.084	3.079	3.082
	ATE	5.370	6.877	4.290
05	RPE	1.952	1.906	1.969
	ATE	18.430	32.851	77.895
06	RPE	2.519	2.518	2.517
	ATE	27.498	38.210	23.064
07	RPE	1.729	1.727	1.728
	ATE	8.941	115.668	27.525
08	RPE	1.948	1.941	1.946
	ATE	92.175	114.930	73.468
09	RPE	2.415	2.402	2.420
	ATE	92.834	64.798	23.353
10	RPE	1.944	1.942	1.934
	ATE	36.709	78.400	38.368
Avg	RPE	2.389	**2.363**	2.366
	ATE	**75.933**	143.169	109.627

parison as [14]. The overall result for the matching score is shown in Table 1. Top matching scores are obtained by EventPointNet on illumination and Superpoint on viewpoint. It is shown that the event data can be advantageous in terms of training valid keypoints under illumination conditions. Additionally, even though qualitative result of Superpoint on Figure 3 seems more dense than the propsed method, quantitative result proves that EventPointNet detects more valid keypoints for feature matching than Superpoint.

4.5 Visual Odometry

To show the robotic application of the method, we make additional experiments in visual odometry based on the 8-points algorithm with the KITTI dataset [3]. The overall experiments are implemented with the functions in the OpenCV framework. Note that the experiments are conducted only with visual local features without other optimization methods such as bundle adjustment. The experiments are proceeded with the methods that guarantee real-time performance, which is a crucial factor for robot applications. Since AKAZE is an accelerated version of KAZE, we use AKAZE for the experiment.

Additionally, to prove that the proposed method is well suited for sudden illumination changes such as tunnels on the road, we modify every second matching frame with gamma correction to darken the image. The descriptor

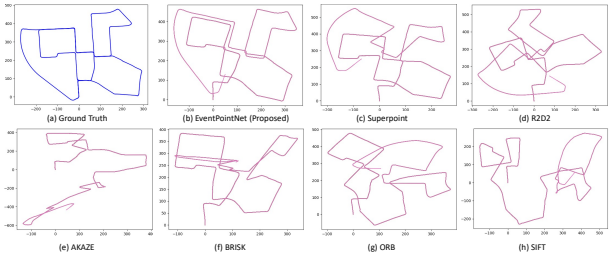

Fig. 4. Visual odometry (VO) results on the KITTI00 dataset.

for keypoints of EventPointNet is the same as the keypoint matching experiments. For other methods, we use descriptors of the each algorithm. The overall procedure is done with the calibrated images processed with the intrinsic camera calibration parameters provided on the dataset.

Figure 4 shows one example of the visual odometry experiment results which is conducted on the KITTI00 dataset. To make the evaluation more plausible, similar to [21], we assume that the scale factor for translation movement is known by the additional sensor such as inertial measurement units (IMU). Although the additional loop closing method is not applied, the proposed method makes less rotation and translation error compared to the other methods. The result trajectories show that the learning-based methods pretend to conserve translation and rotation movement tendency compared with the handcrafted methods. Even though BRISK and ORB perform well on the repeatability test, it is proved that the patch and clamped keypoints are not helpful in estimating camera pose.

The quantitative results with respect to ground truth are shown in Table 2. For evaluating metrics, we adopt absolute trajectory error (ATE) and relative pose error (RPE) used in [22]. We do not evaluate handcrafted methods since most of them fail to estimate trajectories as shown in Figure 4. The results show that EventPointNet makes the lowest ATE in most datasets. The average ATE for EventPointNet is $67.2354\,m$ and $33.6942\,m$ lower than that of Superpoint and R2D2, respectively. In particular, EventPointNet shows a reduction of 30.74% in the absolute trajectory error compared to R2D2. In addition, the results of RPE for all three methods are similar with an average difference of $2.94\,cm$.

5. CONCLUSION

In this paper, a supervised-learning-based keypoint detector trained with neuromorphic camera data named EventPointNet is proposed. An event-based or a neuromorphic camera can capture analogous scenes from diverse conditions, so the data can be meaningful ground-truth data for training a keypoint detector, especially on images with low illumination. Therefore, EventPointNet catches robust and repeatable keypoints regardless of various illumination and viewpoint changes. The exper-

imental results show that the proposed method guarantees high performance in generating local features from any images. Also, the additional experiments on visual odometry show that EventPointNet can be appropriate for robotic applications such as visual SLAM.

ACKNOWLEDGEMENT

This work was supported by the Institute of Information & communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (No. 2019-0-01309, Development of AI Technology for Guidance of a Mobile Robot to its Goal with Uncertain Maps in Indoor/Outdoor Environments)

REFERENCES

[1] J. Hartmann, J. H. Klüssendorff, and E. Maehle, "A comparison of feature descriptors for visual slam," in *Proc. of the European Conference on Mobile Robots (ECMR)*, Barcelona, Sep. 2013.

[2] C. Posch, T. Serrano-Gotarredona, B. Linares-Barranco, and T. Delbruck, "Retinomorphic event-based vision sensors: bioinspired cameras with spiking output," *Proceedings of the IEEE*, vol. 102, no. 10, pp. 1470–1484, Oct. 2014.

[3] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *International Journal of Robotics Research (IJRR)*, vol. 32, no. 11, pp. 1231–1237, Aug. 2013.

[4] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[5] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in *Proc. of the IEEE International Conference on Computer Vision (ICCV)*, Barcelona, Nov. 2011.

[6] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Proc. of the European Conference on Computer Vision (ECCV)*, Graz, May 2006.

[7] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features," in *Proc. of the European Conference on Computer Vision (ECCV)*, Crete, Sep. 2010.

[8] S. Leutenegger, M. Chli, and R. Y. Siegwart, "Brisk: Binary robust invariant scalable keypoints," in *Proc. of the IEEE International Conference on Computer Vision (ICCV)*, Barcelona, Nov. 2011.

[9] P. F. Alcantarilla, A. Bartoli, and A. J. Davison, "Kaze features," in *Proc. of the European Conference on Computer Vision (ECCV)*, Florence, Oct. 2012.

[10] P. Alcantarilla, J. Nuevo, and A. Bartoli, "Fast explicit diffusion for accelerated features in nonlinear scale spaces," in *Proc. of the British Machine Vision Conference (BMVC)*, Bristol, Sep. 2013.

[11] D. DeTone, T. Malisiewicz, and A. Rabinovich, "Superpoint: Self-supervised interest point detection and description," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, Salt Lake City, Jun. 2018.

[12] J. Revaud, C. De Souza, M. Humenberger, and P. Weinzaepfel, "R2d2: Reliable and repeatable detector and descriptor," in *Proc. of Neural Information Processing Systems (NeurIPS)*, Vancouver, Dec. 2019.

[13] Y. Ono, E. Trulls, P. Fua, and K. M. Yi, "Lf-net: learning local features from images," in *Proc. of Neural Information Processing Systems (NeurIPS)*, Montréal, Dec. 2018.

[14] A. Barroso-Laguna, E. Riba, D. Ponsa, and K. Mikolajczyk, "Key. net: Keypoint detection by handcrafted and learned cnn filters," in *Proc. of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Oct. 2019.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[16] A. Z. Zhu, D. Thakur, T. Özaslan, B. Pfrommer, V. Kumar, and K. Daniilidis, "The multivehicle stereo event camera dataset: An event camera dataset for 3d perception," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2032–2039, Jul. 2018.

[17] V. Balntas, K. Lenc, A. Vedaldi, and K. Mikolajczyk, "Hpatches: A benchmark and evaluation of handcrafted and learned local descriptors," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, Jul. 2017.

[18] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors," *International Journal of Computer Vision*, vol. 37, no. 2, pp. 151–172, Jun. 2000.

[19] A. Jakubović and J. Velagić, "Image feature matching and object detection using brute-force matchers," in *Proc. of the International Symposium on Electronics in Marine (ELMAR)*, Zadar, Sep. 2018.

[20] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 27, no. 10, pp. 1615–1630, Aug. 2005.

[21] H. Yu, J. Moon, and B. Lee, "A variational observation model of 3d object for probabilistic semantic slam," in *Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, Montréal, May 2019.

[22] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of rgb-d slam systems," in *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Vilamoura, Oct. 2012.