Detection and Recognition of Traffic Signs from Road Scene Images

Zumra Malik and Imran Siddiqi

Bahria University, Islamabad, Pakistan

zumramalik88@yahoo.com ; imran.siddiqi@bahria.edu.pk

***Abstract*— Automatic detection and recognition of road signs is an important component of automated driver assistance systems contributing to the safety of the drivers, pedestrians and vehicles. Despite significant research, the problem of detecting and recognizing road signs still remains challenging due to varying lighting conditions, complex backgrounds and different viewing angles. We present an effective and efficient method for detection and recognition of traffic signs from images. Detection is carried out by performing color based segmentation followed by application of Hough transform to find circles, triangles or rectangles. Recognition is carried out using three state-of-the-art feature matching techniques, SIFT, SURF and BRISK. The proposed system evaluated on a custom developed dataset reported promising detection and recognition results. A comparative analysis of the three descriptors reveal that while SIFT achieves the best recognition rates, BRISK is the most efficient of the three descriptors in terms of computation time.**

***Index Terms*—Road sign detection, Road sign recognition, Hough transform, SIFT, SURF, BRISK.**

I. INTRODUCTION

The last two decades have witnessed a tremendous increase in the number of road vehicles all over the world thanks to the technological advancements in the motor industry and the availability of vehicles at economized rates. With this remarkable growth of road traffic, the number of accidents has also increased significantly. Among different causes of these accidents, a major cause is the ignorance of road signs by drivers. Developing automated systems to assist the road drivers by detecting and recognizing the road signs and alerting the drivers could possibly serve to reduce the number of accidents. This automatic detection and recognition of road signs from real world scenes is the subject of our study. The main challenges arise due to variations in lighting and weather conditions, low resolution of captured images and similarity among different signs.

The problem of detection and recognition of road signs has been an active area of research for many years. Traditionally, detection is carried out by using the shape and color information of the signs while recognition is implemented using shape matching strategies. Most of the detection schemes proposed in the literature exploit the color properties of the signs and rely on color based thresholding [1-5] to segment the sign from rest of the image. In addition, distance transforms [6], projection features [7,8] and Hough transform [5] have also been investigated to detect the road signs based on their shapes. Methods based on weighted mean shift algorithm [9] and genetic algorithms [10] have also been proposed. For recognition, typically, each road sign is represented by a set of features and the features of query sign are matched with the stored (feature) templates to determine the sign class. In case of supervised approaches, features extracted from different sign classes are fed to a learning algorithm which learns to discriminate between these classes. Classifiers explored for recognition of road signs include neural networks [1, 11, 12], support vector machine [4, 13, 14], hidden Markov models [7] and AdaBoost [15, 16]. Despite these and other significant research contributions, the problem of detection and recognition of traffic signs remains challenging and open to research due to the complexities it offers.

This paper presents an effective method for detection and recognition of traffic signs from images of road scenes. The proposed detection scheme exploits the color and shape attributes of the signs and is based on color based segmentation and region of interest extraction using the Hough transform. For recognition, we present a comparative analysis of three well-known shape descriptors namely SIFT, SURF and BRISK. We discuss the proposed detection and recognition methodology in sections II and III of the paper.

978-1-4799-7505-1/14 $31.00 © 2014 IEEE
DOI 10.1109/FIT.2014.68

Experimental results and the accompanying analysis are presented in Section IV while the last section concludes the paper with a discussion on potential future research directions on this problem.

II. DETECTION OF ROAD SIGNS

Detection of signs from road images includes color based segmentation followed by shape analysis and application of geometrical constraints to segment the sign from rest of the image. These steps are detailed in the following subsections.

A. Color based Segmentation

Road signs appear in different shapes and colors. Typical colors include red, blue, green and yellow. Red circular signs are the warning symbols while red triangular signs represent the mandatory signs. Blue and yellow signs generally appear as rectangles and represent informative signs and construction warnings respectively. Green signs are common on motorways and signal the milestones and distances. The different categories of road signs are summarized in Table 1.

Table 1: Summary of road sign categories

Color	Shape	Category
Red	Circular	Warning signs
Red	Triangular	Mandatory signs
Green	Rectangular	Directions
Blue	Rectangular	Informatory signs
Yellow	Rectangular	Construction/road works

To detect potential road signs from images, we exploit the color information of the signs to segment possible regions of interest from rest of the image. Prior to applying a color based thresholding, we first convert the image from RGB to the HSV color space. The H component in the HSV color scheme represents the chromatic information; the S component corresponds to different shades of a particular color while the V component specifies the brightness of a color. The motivation of using HSV color space lies in the fact that it is one of those color spaces which distinguish color from intensity making it non-sensitive to variations in the illumination, a desirable attribute in our detection system. An example of images converted from RGB to HSV color space is illustrated in Figure 1.

(a) (b)

Figure 1: Examples of road scene images (a) RGB color space (b) HSV color space

Once the image is converted to HSV color space, we need to find appropriate thresholds which allow segmentation of our colors of interest. The thresholds must be relaxed enough to allow capturing the variations within the same color and strict enough not to include a large number of unwanted objects. Naturally, for each of the red, green, blue and yellow signs, a different set of threshold values are required on the hue, saturation and value components. In order to compute these thresholds, for each of the four colors of interest, we extract the hue, saturation and value components from a set of sample road sign images corresponding to different lighting conditions. A histogram of each color component for different sign images is computed to determine the thresholds. The threshold values chosen by analyzing the H, S and V histograms for each of the red, blue, green and yellow color are summarized in Table 2 and Table 3. For red color, two different thresholds on the hue component are listed corresponding to the time of image capture (day or evening). These thresholds are consistent with the ones computed in other similar studies [1-5].

Table 2: Threshold values to segment red color

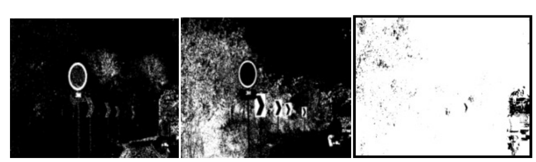
Component	Day time	Evening time
Hue	$0.97<H<1$	$0<H<0.06$
Saturation	$0.5<S<1$	$0.5<S<1$
Value	$0.2<V<1$	$0.2<V<1$

Table 3: Threshold values to segment green, blue and yellow colors

Color	Component	Day/Evening time
Green	Hue	$0.40<H<0.50$
	Saturation	$0.7<S<1$
	Value	$0.5<V<1$
Blue	Hue	$0.33<H<0.60$
	Saturation	$0.7<S<1$
	Value	$0.3<V<1$
Yellow	Hue	$0.10<H<0.20$
	Saturation	$0.9<S<1$
	Value	$0.8<V<1$

Given an input image, the pixels satisfying the sets of threshold values listed in Tables 2 and 3 are considered as potential road sign regions. A binary image is generated in which all pixels in the specified threshold ranges are set to 1 and all the remaining pixels are set to 0. The result of applying (red) color segmentation on a sample image is illustrated in Figure 2 while Figure 3 demonstrates the thresholding of the same image on individual color components.

Figure 2: A road scene image and result of (red) color based thresholding



(a) (b) (c)

Figure 3: Thresholded image (a) Hue (b) Saturation (c) Value

Once the potential road sign regions are segmented using color information, we analyze the shapes of the detected regions to eliminate false positives and detect the true signs.

B. Shape Analysis

Color based thresholding serves as a filter to eliminate the image regions which are not likely to contain road signs. Naturally, the image may have other objects of the same colors as road signs and further processing is required to segment signs from rest of the image. For this purpose, we exploit the shape attributes of the road signs. The output of color based segmentation is a binary image with potential sign regions in white and the background in black. Morphological dilation with a squared structuring element is applied to the binary image to fill small gaps and missing boundaries.

We next carry out connected component labeling on the dilated image to find the number of objects in the image and their properties. All road signs are circular, triangular or rectangular and obey certain constraints on the aspect ratio of their bounding box. In our implementation, we compute the aspect ratio of each connected component and eliminate all components with an aspect ratio greater than an empirically determined threshold.

To distinguish between road signs and other objects in the image, we employ the Hough transform. Hough transform is the most well-known method for detection of boundaries or curves in an image that transforms the points in Cartesian space to a parameter space allowing detection of lines and other geometrical shapes [17, 18]. In our implementation, we apply the Hough transform to detect circles and lines in the image. The minimum and maximum radii values to the circular Hough transform are set to 5% and 25% of the image height respectively. If a circle or a set of connected lines (4 in case of rectangular sign and 3 in case of triangle sign) is detected in the image, the component is retained as a valid road sign, otherwise, it is discarded. The detected region is then extracted from the original image as a road sign. In the next section, we discuss the details of the recognition of detected road signs.

III. RECOGNITION OF ROAD SIGNS

Recognition of road signs is carried out using three widely used descriptors, SIFT, SURF and BRISK. An overview of each of these is presented in the following.

A. Scale Invariant Feature Transform (SIFT)

SIFT is a powerful computer vision algorithm that combines feature detector and feature descriptor. SIFT features are invariant to scale, rotation and to some extent view point as well. The feature detector determines key points in the image and associates an orientation with each key point. A descriptor is then computed for each key point.

The scale invariance in SIFT is achieved by generating a scale space using Difference of Gaussians (DoG). Difference of Gaussians is obtained as the difference of Gaussian blurring of an image with two different values of σ (σ and $k\sigma$). This process is repeated for different octaves of the image in the Gaussian pyramid. Once the DoGs are computed, the images are searched for local extrema. For each pixel in the image at a given scale, its value is compared with 8 of its neighbors on the same scale, 9 pixels in the next scale and 9 pixels in the previous scale. The pixel is chosen as a potential key point if it is a local extrema. These potential key points are then filtered to remove the points with low contrast and those lying on edges and the remaining points are chosen as the final key points. Once the key points are determined, an orientation is assigned to each key point based on the magnitudes and directions of gradients in the neighborhood of each key point. To generate the descriptor for a key point, the 16x16 neighborhood around the key point is divided into small blocks of size 4x4. For each block, an eight bin orientation histogram is computed and all the 16 histograms are concatenated to form a 128 dimensional vector – the SIFT descriptor. Two images are matched by matching their key points through identification of their nearest neighbors. In order to avoid false matches, if the ratio of closest-distance to the second closest distance is greater than a threshold (0.8), the key points are rejected [19].

B. Speeded up Robust Features (SURF)

SIFT, although the most widely used and accepted descriptor for feature matching suffers from the drawback of high computational cost. In 2006, H. Bay et al. [20] proposed a speeded-up variant of SIFT which was termed as SURF (Speeded-up Robust Features). While SIFT computes a histogram of local gradients in the neighborhood of each key point giving a 128 dimensional descriptor, SIFT relies on sums of Haar wavelet components producing a 64 dimensional feature descriptor.

The Laplacian of Gaussian (LoG) in SIFT is approximated by Difference of Gaussian to generate the scale space. SURF, on the other hand, approximates LoG with a box filter. The significant reduction in computation time is gained by using integral images which allow a very fast computation of box type filters. The detection of key points is based on (the determinant of) Hessian matrix.

The determinant of the Hessian matrix at each location in the image over different scales is computed and stored to search for local maxima. Once the interest points are located, an orientation is associated with each key point (similar to SIFT) to allow rotation invariance. The orientation is determined by computing the Haar wavelet response (in horizontal and vertical directions) in a circular neighborhood of each key point. The wavelet responses are plotted as points with the strength of horizontal and vertical responses on the two axes. A sliding window of 60 degrees is then used to sum up all responses within the window and the dominant orientation is selected as the orientation of the key point.

The SURF descriptors are calculated by considering a square region (of size 20) around each key point oriented along its respective orientation. The region around the key point is divided into sub-regions of size 4 x 4 and for each block, Haar wavelet response in horizontal and vertical directions is computed. Each sub-region is then represented by a four dimensional descriptor which includes the sum of wavelet responses (d_x and d_y) and the sum of the absolute values of d_x and d_y as summarized in Eq. 1.

$$v = \left(\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|\right) \tag{1}$$

These descriptors are computed for each of the sub regions giving a 64 dimensional descriptor for each key point.

C. Binary Robust Invariant Scalable Key points (BRISK)

BRISK [21] employs FAST [22, 23] (Features from Accelerated Segment Test) score as a measure for saliency. The FAST algorithm considers the intensity of pixels on a circle around the candidate pixel *C*. The pixel *C* is considered a FAST corner if a set of *N* contiguous pixels on the circle are either all brighter or all darker than the intensity of the pixel *C* (plus some threshold). Typically, the FAST mask tests for 9 contiguous pixels in a 16-pixel circle.

BRISK also employs 9-16 masks which are applied to each octave and intra-octave to find the points of interest. Non-maxima suppression is then applied to the points satisfying the FAST criterion. To formulate the BRISK descriptor, a sampling pattern is defined around each key point which, in general, comprises concentric rings.

In order to compute the orientation of the key point and the associated descriptor, a distinction is made between short pairs and long pairs. Short (long) pairs comprise all pairs of sampling points such that their distance is below (above) a certain threshold. BRISK employs the long pairs of sampling points to determine the key point orientation while the short pairs are used to produce the descriptor.

Orientation of key point is computed by using local gradients between the (long) sampling pairs. The descriptor is computed by comparing the intensity values of the short pairs. For each pair, if the intensity of the first point is larger than that of the second point, the corresponding bit of the descriptor is set to 1. Otherwise, it is assigned a 0.

$$b = \begin{cases} 1, I(p_j,\sigma_j) > I(p_i,\sigma_i) \\ 0, otherwise \end{cases} \tag{2}$$

The typical sampling pattern used for BRISK and the recommended threshold values to find short and long pairs generate a 512 bit sequence (descriptor) for each key point. Two descriptors are compared using the Hamming distance which is simply the sum of XOR operation between them.

After having presented the three descriptors, we discuss the experimental results in the next section.

IV. EXPERIMENTS AND RESULTS

This section presents in detail the series of experiments carried out to validate the proposed methodology for detection and recognition of road signs. We first discuss the data set used in our experiments followed by the detection and recognition results.

A. Database

We have employed a custom developed database where road sign images are captured by a camera mounted on a moving vehicle. The images were taken on multiple days at different times of the day (morning, afternoon and evening) while night time images are beyond the scope of our study. The database comprises a total of 32 different road signs with 5-6 images per road sign giving a total of 172 images. The road signs considered in our study are summarized in Table 4.

Table 4: Road signs considered in our study

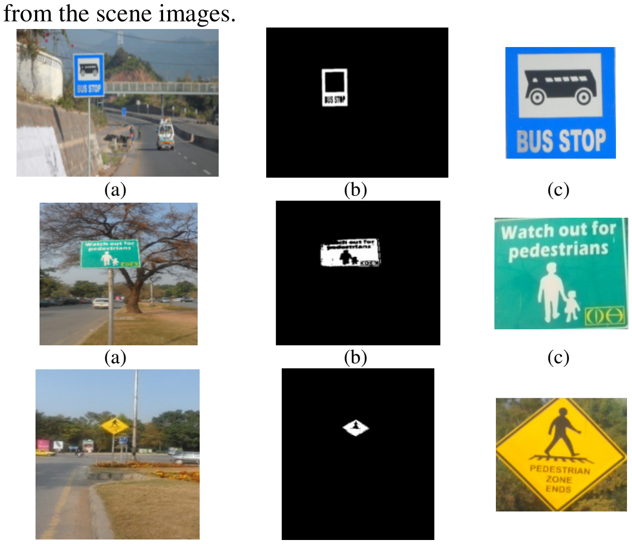
B. Detection Results

Detection is carried out using color based segmentation and shape analysis as discussed earlier. A total of 169 road signs images are correctly detected realizing an accuracy of 98.25%. The category wise detection results are summarized in Table 5.

Table 5: Detection performance

Sign Color	Red	Blue	Green	Yellow	Total
Total Signs	116	28	6	21	172
Detected	114	28	6	21	169

Figure 4 illustrates some examples of the road signs detected from the scene images.

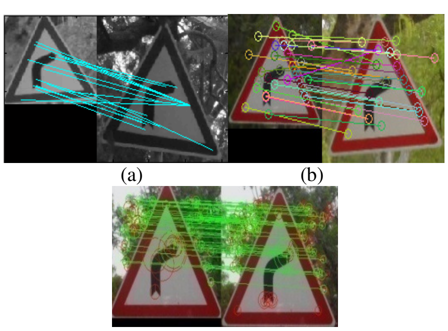


(a) (b) (c)

Figure 4: Detection results (a) Original Images (b) Color based segmentation and shape analysis (c) Detected road signs

C. Recognition Results

Recognition results are reported for each of the descriptors SIFT, SURF and BRISK. In the first series of experiments, one image of each class is used as test and the remaining images (4 or 5) of each class are used in training. Figure 5 illustrates the matching of an example road sign using the three descriptors while Table 6 summarizes the recognition rates. It should be noted that recognition is carried out in two different scenarios. In the first scenario, the manually segmented query road sign is compared with the signs in the database to determine the recognition performance. In the second scenario, the output of the detection system (which may not be a perfectly segmented road sign) is directly compared with the signs in the database to match the real world situations.



(a) (b) (c)

Figure 5: Matching using (a) SIFT descriptor (b) SURF descriptor (c) BRISK descriptor

Table 6: Recognition rates on three descriptors

	Recognition Rate	
Feature	**Scenario I**	**Scenario II**
SIFT	100%	93.75%
SURF	93.75%	81.25%
BRISK	93.75%	87.5%

It can be observed from Table 6 that SIFT outperforms SURF and BRISK in both the evaluation scenarios. When manually segmented road signs are matched, SIFT recognizes all of the signs correctly. For more realistic scenario of feeding the output of detection module to the recognition module, SIFT still performs better realizing a recognition rate of 93.75% while SURF and BRISK report recognition rates of 81.25% and 87.5% respectively. The drop in recognition rates in scenario II is due to the fact the output of detection is not always a perfectly segmented road signs and may also contain parts of the background scene. In order to study the impact of number of training samples on the overall recognition rates of the three descriptors we carry out a series of evaluations by varying the training samples from 1 to 4 for each of the road signs. The results of these evaluations can be seen in Figure 6. It can be noticed from the results in Figure 6 that SIFT is the least sensitive to the number of training examples and maintains relatively stable recognition rates with the change in number of training samples. BRISK features seem to be the most sensitive to number of training samples where the recognition rates drop from 93.7% to 84.3% when reducing the training samples from 4 to 1.

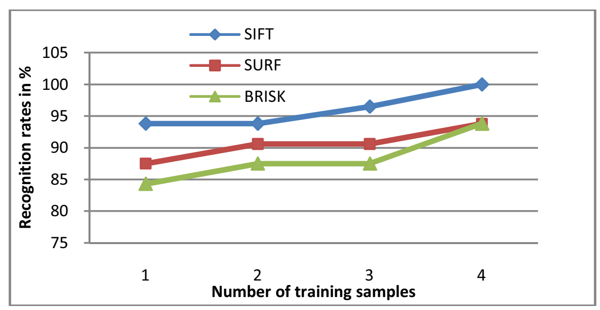


Figure 6: Recognition rates on three descriptors as a function of number of training samples

Since the main idea of BRISK descriptors is to speed up the matching process in comparison to SURF and SIFT, we also carry out a comparison of the execution time of the three descriptors. The execution times are computed by matching each road sign with all the training images and taking an average. Table 7 summarizes the (average) relative execution times of the three features.

Table 7: Relative matching times of three descriptors

Feature	**Relative matching time**
SIFT	2.1t
SURF	1.4t
BRISK	t

It can be seen that BRISK is almost twice as fast as SIFT making it an appropriate descriptor for a real time matching problem like recognition of road signs. The present system is developed in Matlab and the average value of *t* on an Intel i3 machine with 4GB RAM is 4.3 seconds. Naturally, for real time implementation, BRISK could be implemented on an appropriate platform from the view point of an application.

V. CONCLUSION

This study presented an effective and efficient road sign detection and recognition methodology which is invariant to changes in illumination, scale and viewing angle. Detection is based on color segmentation carried out in the HSV color space followed by a shape analysis of candidate regions using the Hough transform. Recognition is carried out using three state-of-the-art feature matching algorithms namely SIFT, SURF and BRISK. These descriptors have been effectively applied to a number of computer vision problems and we investigated their effectiveness on recognition of road signs. Experiments carried out on a custom developed database revealed that SIFT outperforms SURF and BRISK in terms of recognition rate whereas BRISK turns out to be the most efficient in terms of computation time while maintaining acceptable recognition rates.

Presently, the system works on day time images only. It could be further enhanced to detect and recognize road signs from night time images as well. Although color based segmentation provides acceptable detections, it would be interesting to investigate other, more sophisticated techniques which may include supervised detection methods. In addition, it would also be interesting to perform detection on video rather than still images. Since neighboring frames in a video exhibit temporal redundancy, information from multiple frames could be employed to effectively detect the signs resulting in improved detection rates. The detection of signs and their subsequent recognition can also be complemented by a text based output to alert the drivers. The system may also be extended to include detection and recognition of textual information in the form of directions and milestones etc.

REFERENCES

1. De La Escalera, A., Moreno, L. E., Salichs, M. A., & Armingol, J. M., Road traffic sign detection and classification. IEEE Transactions on Industrial Electronics, 44(6), 848-859.
2. Bénallal, M., & Meunier, J., *Real-time color segmentation of road signs*. In Proc. of Canadian Conference on Electrical and Computer Engineering, 2003, pp. 1823-1826.
3. Broggi, A., Cerri, P., Medici, P., Porta, P. P., & Ghisio, G., *Real time road signs recognition*. In IEEE Intelligent Vehicles Symposium, 2007, pp. 981-986.
4. Maldonado-Bascon, S., Lafuente-Arroyo, S., Siegmann, P., Gomez-Moreno, H., & Acevedo-Rodriguez, F. J., *Traffic sign recognition system for inventory purposes.* In Intelligent Vehicles Symposium, 2008, pp. 590-595.
5. Huang, H., Chen, C., Jia, Y., & Tang, S., *Automatic detection and recognition of circular road sign*. In Proc. of International Conference on Mechtronic and Embedded Systems and Applications, 2008, pp. 626-630.
6. Gavrila, D. M., *Traffic sign recognition revisited*. In Mustererkennung 1999, pp. 86-93.
7. Hsien, J. C., Liou, Y. S., & Chen, S. Y., *Road sign detection and recognition using hidden Markov model. Asian Journal of Health and Information Sciences*, 2006, *1*(1), pp. 85-100.
8. Gracia, M.A., Sotelo M. A., Gorostiza, E. M., *Traffic Sign Detection in Static Images using Matlab.* In Procc. of Emerging Technologies and Factory Automation, 2003, pp. 212-215.
9. Ruta, A., Porikli, F., Watanabe, S., & Li, Y., *In-vehicle camera traffic sign detection and recognition. Machine Vision and Applications*, 2011, *22*(2), pp. 359-375.
10. De La Escalera, A., Armingol, J. M., & Salichs, M. A., *Traffic sign detection for driver support systems*. In International conference on field and service robotics, 2001.
11. Kim, G. H., Sohn, H. G., & Song, Y. S., *Road Infrastructure Data Acquisition Using a Vehicle- Based Mobile Mapping System.* Computer-Aided Civil and Infrastructure Engineering, 2006, 21(5), pp. 346-356.
12. Ohara, H., Nishikawa I., Miki, S. and Yabuki, N., *Detection and recognition of road signs using simple layered neural networks*. In Proc. of the 9th Int. Conf. on Neural Information Processing. 2002.
13. Fard, A.L., Chalechale, A. and Ebrahimi, S.A., *Recognition of the Triangular and Rectangular Road Signs Using Passing Lines Inersection.* 16th Intl. Symposium on Artificial Intelligence and Signal Processing, 2012: pp. 310- 315.
14. Maldonado-Bascon, S., Lafuente-Arroyo, S., Gil-Jimenez, P., Gomez-Moreno, H. and Lopez-Ferreras, F., *Road-Sign Detection and Recognition Based on Support Vector Machines.* IEEE Transactions on Intelligent Systems, 2007. 8(2): p. 264- 278.
15. Bahlmann, C., Zhu, Y., Ramesh, V., Pellkofer, M., & Koehler, T.. *A system for traffic sign detection, tracking, and recognition using color, shape, and motion information*. In Proc. of Intelligent Vehicles Symposium. 2005.
16. Chen, S. Y., & Hsieh, J. W.. *Boosted road sign detection and recognition.* In Proc. of Intl. Conference on Machine Learning and Cybernetics, 2008. pp. 3823–3826.
17. Illingworth, J., & Kittler, J. *A Survey of the Hough Transform*, Computer vision, graphics, and image processing, 44(1), 87-116.
18. Rizon, M., Haniza, Y., Puteh, S., Yeon, A., Shakaff, M., Abdul Rahman, S., and Karthigayan, M., *Object detection using circular Hough transform*, 2005.
19. Lowe, D. G., *Object recognition from local scale-invariant features*. Proc. of the 7th Intl. conference on Computer Vision, 1999.
20. Bay, H., Tuytelaars, T., & Van Gool, L. *Surf: Speeded up robust features*. In Computer Vision–ECCV 2006, pp. 404-417.
21. Leutenegger, S., Chli, M., & Siegwart, R. Y., *BRISK: Binary robust invariant scalable keypoints*. Proc. Of Intl. conference on Computer Vision, 2011, pp. 2548-2555.
22. Rosten, E., & Drummond, T., *Fusing points and lines for high performance tracking*. Proc. Of International Conference on Computer Vision, 2005, pp. 1508-1515.
23. Rosten, E., Porter, R., & Drummond, T., *Faster and better: A machine learning approach to corner detection*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(1), 105-119.