https://doi.org/10.2352/ISSN.2470-1173.2017.9.IRIACV-258

Efficient visual loop closure detection in different times of day

Can Erhan[1], *Evangelos Sariyanidi*[2], *Onur Sencan*[1], *Hakan Temeltas*[1]

[1] *Control and Automation Engineering Department, Istanbul Technical University, Istanbul, Turkey.*
[2] *School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK.*

Abstract

Performing reliable and computationally efficient loop closure detection in real-world environments still remains a challenging problem. In this paper, we propose a novel method for efficient loop closure detection in different times of day. An illumination invariant color transform is applied to images that are represented by a whole-image descriptor, named PALM. The efficiency of our method resides either in description of the places or in image matching in which FLANN is used for fast nearest neighbor search. With this approach, searching time is decreased about 70 times compared to standard brute-force search with no significant loss of accuracy. According to the experiments that are performed in real-world datasets, the proposed method successfully accomplishes to detect loops under varied illumination conditions with high accuracy, and it allows real-time operation for long-life localization and mapping.

Introduction

In the context of robotic navigation, constructing a map of an unknown environment and localizing the agent itself in it are essential tasks to accomplish the missions. Although both tasks initially appear to be independent, they are closely related and considered as a single problem, known as SLAM (*Simultaneous Localization and Mapping*). Loop closing is defined as the correct identification of previously visited place in terms of SLAM. This ability is crucial for not only accurate localization, but also creating consistent maps by minimizing the accumulated errors arising from the sensory information.

Traditional SLAM approaches have relied on range sensors such as LIDARs in terrestrial environments or SONARs in underwater. However, the usage of vision-based sensors have been quite popular in recent years due to their competitive prices and compact structure being able to provide rich information. When vision is the source, loop closure detection is performed by comparing the images directly [1].

In order to generate consistent maps, avoiding false detections is a key factor must be taken into consideration [2, 3]. Perceiving different places as the same, known as perceptual aliasing, is one of the main problems of visual loop closure detection. Similarly, perceiving same places as different is another problem that must be dealt with. For instance, the place in the first visit and the one after the second visit may differ due to changed illumination. In real-time operations, detecting loop closures can be computationally expensive with complex image processing techniques. Various loop closing systems suffers from this problem in spite of their high performance [2]. Therefore, it is essential to use a fast and efficient method even for the agent traverses a large map.

The rest of the paper is organized as follows: We fist review the related works, and then we introduce the efficient loop closure methodology. In the following section, the proposed method is validated in real-world datasets. Finally, the last section is dedicated for conclusion and future works.

Related Work

Visual loop closure detection is still an active and challenging problem despite the extensive studies investigated for many years. This problem becomes even more challenging when it needs to be solved for long sequences.

The most popular approach for localization is bag-of-words (BoW) models [4], which is employed by the well-known FAB-MAP [2]. While BoW models are suitable for large-scale operations, they have a large computational storage and time cost because of the keypoint detection that is performed in the first place. In [1], a tree structure was employed to efficiently identify correct loop closures. It was also reported that the indexing structure for fast retrieval of loop closure candidates shows superior performance to BoW on a moderate size dataset.

Another approach to loop closure detection is to use sequential frames [5, 6]. Instead of matching a single frame, the method in [5] calculates the best candidate matching frames within every local sequences. With this approach, place recognition was performed in real-world data that have extreme perceptual change such as sunny days in summer and stormy winter nights. Inspired by this work, in [6], sequential binary codes of the images was used for long-life localization up to a 3000 km long trajectory. The image matching was also performed efficiently with FLANN.

Several works have particularly focused on the loop closure problem under different lighting conditions [5, 7, 8, 9]. To handle intense illumination (e.g. full of dark and quite dim environments), the descriptors of both gray and depth images were combined in [9]. It was also presented that using depth image in full of dark places might be useful for image matching.

In visual loop closure detection, the description method is the key point to achieve good matching results. In this paper, we employ PALM descriptor [10], which has been introduced by the authors of this paper. It has been shown that PALM outperforms the state-of-the-art description methods such as WI-SIFT [11], BRIEF-Gist [3] and LDB [12] in terms of loop closing accuracy. Furthermore, we show that using illumination invariant color transform before the descriptor is computed improves loop closure performance significantly, especially in intense illumination changes over images. FLANN based search is used for efficient matching of images. Compared to standard brute-force search, we show that FLANN can be useful for loop closure detection by decreasing computational overhead up to 70 times, which is important for large-scale localization and mapping.

Efficient Loop Closure Detection

In this section, we first explain the description method that is able to describe the scene image as a whole. Illumination invariant color transform is also emphasized. At last, we express the image matching approach to detect loop closures efficiently.

Image description

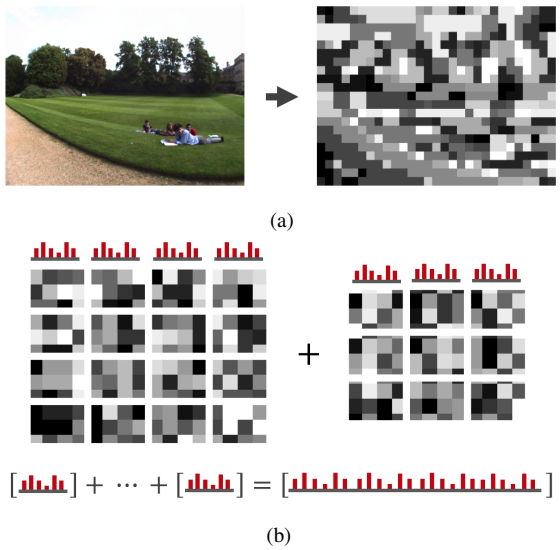
PALM (*Patterns of Approximated Localized Moments*) [10] is a histogram based global image descriptor that allows reliable and real-time loop closure detection in the presence of strong perceptual aliasing. This method relies on computing local Zernike moments (ZMs) [13, 14] across the image. When used locally, Zernike moments provide a highly discriminative property, which is an essential quality to deal with perceptual aliasing (i.e. to discriminate places that are visually very similar). PALM is computed in two stages as illustrated in Fig. 1: (a) extracting the pattern image and (b) constructing the descriptor vector.

Firstly, ZMs are computed for $k \times k$ sized local patches across the image. Those patches are selected sequentially to start from the top-left of the image, and then move rightwards by a step of s pixels with the constraint of $k = cs$ where c indicates overlap density. Let I_{ij} be a local patch that yields a set of complex coefficients according to the moment order n as follows:

$$\mathbf{Z}_n(I) = \{Z_1^1, Z_2^2, Z_3^1, \ldots, Z_n^m\}. \tag{1}$$

Since a large number of coefficients are obtained in total, quantization is applied by taking only their signs to compress the data. Those binary values are combined into an integer which resides in a single pixel of the pattern image as depicted in Fig. 1(a). Furthermore, approximating Zernike bases and employing integral images in the computation of moments reduces the computational overhead tremendously without any loss of matching accuracy.

Reshaping the pattern image as a descriptor vector is not practical since loop closure pairs may contain spatial inconsistencies. To this end, pattern image is partitioned to $g \times g$ subregions, and each subregion is described with a local histogram. An inside



(a)

(b)

Figure 1: Illustration of computing PALM descriptor: (a) pattern image extraction and (b) descriptor vector construction.

partitioning is also done with $(g-1) \times (g-1)$ subregions which overlaps with the ones in complete partitioning. Those local histograms build for each subregions are concatenated to construct the final descriptor vector as illustrated in Fig. 1(b).

Illumination invariant color transform

A typical problem that needs to be addressed in loop closure detection is perceiving same places as different. When a loop closure takes place, the image collected in the first visit and another in the second visit may differ due to the illumination. If we try to match those images, they may not be matched. To improve robustness against illumination, we apply an illumination invariant color transform to the images. Similar strategies are employed in [6, 8, 15].

Let $\{r, g, b\}$ be the color components of a 3-channel floating-point color image which is mapped to the range of $[0, 1]$. A single channel illumination invariant image, $\mathbb{I}$, is computed as:

$$\mathbb{I} = 0.5 + \log(g) - \alpha \log(b) - (1-\alpha)\log(r) \tag{2}$$

where α is a parameter which satisfies the following constraint:

$$\frac{1}{\lambda_g} - \frac{\alpha}{\lambda_b} - \frac{1-\alpha}{\lambda_r} = 0 \tag{3}$$

where λ_r, λ_g and λ_b denote the peak spectral responses of corresponding color channels. Those values can be found in the datasheets of cameras. Fig. 2 gives examples images and their illumination invariant color transformed versions for $\alpha = 0.4$.

Image Matching

We formulate the loop closure detection as an image matching problem via nearest neighbor search (NNS). The aim is to find the image that closes the loop (i.e. the most similar one) to the most recent image in the images collected throughout the trajectory. Let $D = \{d_1, d_2, d_3, \ldots, d_{t-1}\}$ be the set of descriptor vectors (i.e. points in a metric space S) that are computed from the images previously seen, and $d_t \in S$ denotes the descriptor of the most recent image. Linear or brute-force search may help to find the nearest neighbor, but it is not enough efficient to be operated for long-life localization because of the growing size of the set D. Brute-force search is also quite time consuming for image matching [1].

To efficiently match the images in a given set, FLANN (*Fast Library for Approximate Nearest Neighbors*) [16, 17] is employed to perform fast approximate nearest neighbor search. However, there is a trade-off between efficiency and accuracy. In approximate search, the set D must be preprocessed or indexed in such a way that the most recent one d_t can be found rapidly in that set. The distance metric we use during this search is l_1-*norm*, which is one of the most simple metric available. A loop closure hypothesis is accepted if the distance between the matched frames is below a threshold τ since there may not be any loop closure.

Experiments

In this section, several experiments are conducted to evaluate the proposed method. First, we explain the experimental setup including datasets and methodology of the experiments. Then we present the results of the proposed method in terms of loop closing performance.

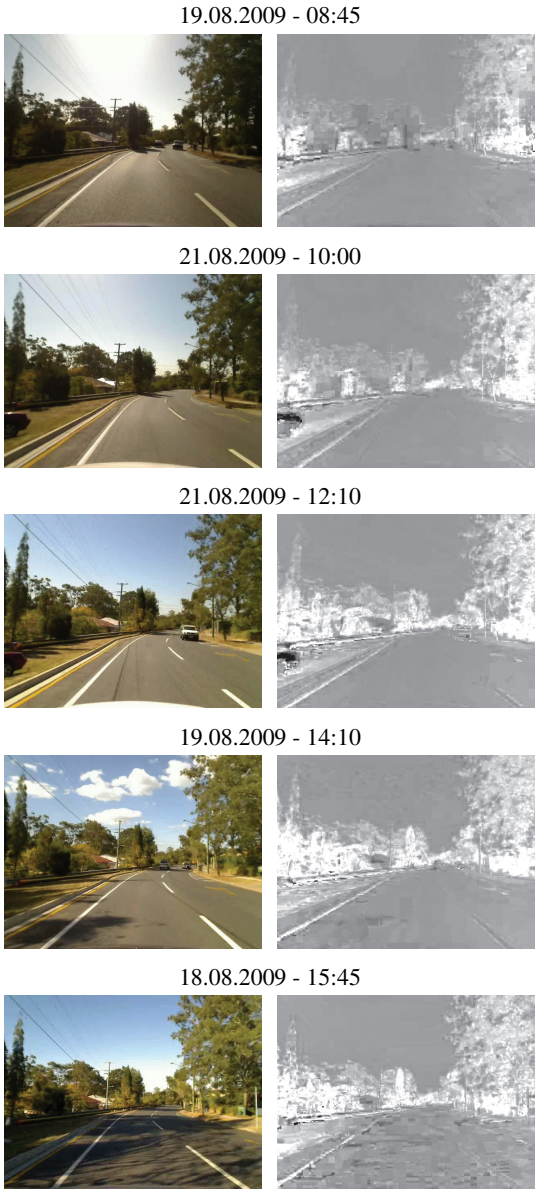


Figure 2: (a) Example images taken from the same position for the first group of StLucia dataset and (b) their illumination invariant color transformed versions for $\alpha = 0.4$.

Experimental Setup

StLucia dataset [7] has 10 sequences of 640×480 sized highly compressed images in a video recorded in 15 fps. Each sequence is about 12 km long and traversed by a car 10 times at different day times. First group (i.e. the first 5 sequences) are collected before 3 weeks than the other group, which contains the same hours as the first one. We employ only the first group of which recording times and frame counts are listed in Table 1. GPS positions are provided. These sequences contain a highly varied range of terrain and scenery, including intense illumination changes and shadowing effects. Fig. 2 illustrates example images that are taken from the same position for the first group.

Table 1: Recording times and frame counts of the first group in StLucia dataset.

Recording Date & Time	Frame Count
19.08.2009 - 08:45	21815
21.08.2009 - 10:00	20457
21.08.2009 - 12:10	18077
19.08.2009 - 14:10	21373
18.08.2009 - 15:45	21433

We use OpenCV implementation of FLANN, which is a wrapper library of the original source-code provided by [16, 17]. The sequence in the middle (i.e. the one recorded at 12:10) is used to create FLANN index satisfying 95% precision of returning the exact nearest neighbor. Besides, we use the source-code of PALM provided by [10]. The default settings (i.e. $k = 32$, $g = 5$, $c = 4$, $n = 2$) of PALM is used as suggested. Lastly, illumination invariant color transform is applied to each image before the descriptor is computed. Since camera specs are not provided with the dataset, we set $\alpha = 0.4$ based on the exemplar α values stated in [8].

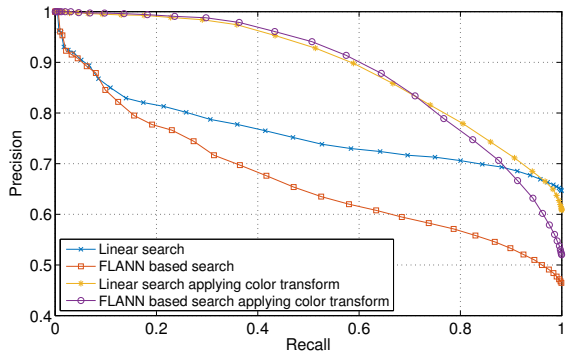
We evaluate the proposed method by using the metrics that are standard in loop closure literature: *precision* is the ratio between true positives and the total amount of loop closures detected, and *recall* indicates the percentage of true positives in the total amount of loop closures available on the entire sequence. Every detection is considered as true positive if the locations of matched frames are close as up to 20 m. We also report precision-recall curves, which are obtained by sweeping the matching threshold τ from 0 to the value that makes recall 100%.

Results

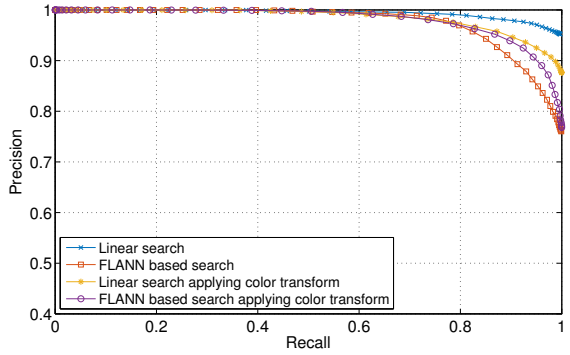
As mentioned above, the sequence recorded at 12:10 is used for indexing since illumination is more stable than the others (i.e. the ones recorded at 08:45, 10:00, 14:10 and 15:45) as illustrated in Fig. 2. Accordingly, we perform four separate experiments for each dataset excluding the one used for indexing.

Fig. 3 presents the precision-recall curves not only with FLANN based search but also with linear search (i.e. brute-force search). The effect of illumination invariant color transform is also considered. As it is seen, the proposed method performs better in the sequences recorded at 10:00 and 14:10 since the illumination is much more similar to the indexed sequence. However, applying color transform has slightly negative effect on those sequences in contrast to the significant improvement in the ones at 08:45 and 15:45. Consequently, it is obvious that applying illumination invariant color transform gives satisfactory results in terms of loop closure performance.

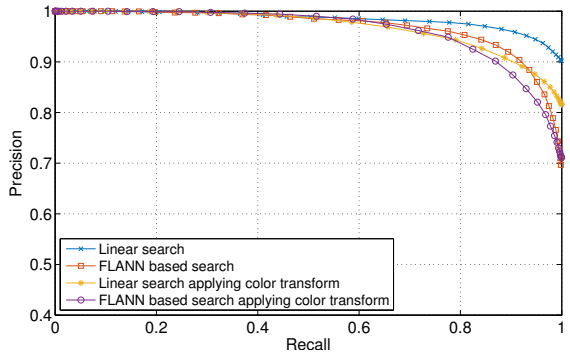
As Fig. 3(a) and Fig. 3(d) depict, FLANN based search is almost as successful as the linear search when color transform is applied. Note that the FLANN index is created to satisfy 95% precision since there is no algorithm that performs considerably better than linear search [16, 17]. Table 2 shows average processing times of the separate processes contained by the proposed method. Those values are computed with standard laptop computer through $\sim$20 k frames which corresponds to 12 km long trajectory. As it can be seen, image matching with FLANN is about 70 times faster than linear search. With that dramatic difference, FLANN is quite acceptable for long-life localization even its performance slightly lower than linear search.



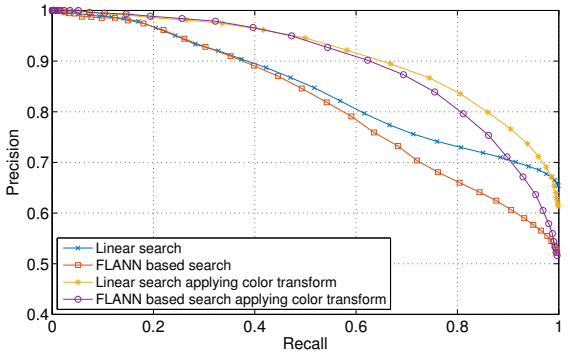
(a) 12:10 vs 08:45



(b) 12:10 vs 10:00



(c) 12:10 vs 14:10



(d) 12:10 vs 15:45

Figure 3: Precision-recall curves for comparing matching performance. Illumination invariant color transform is also considered.

Table 2: Average processing times of separate processes contained by the proposed method.

Process	Avg. Time (ms)
Applying color transformation	5.9203
Extracting PALM descriptor	1.3207
Image matching with linear search	9.1831
Image matching with FLANN	0.1321
Complete system with linear search	16.4241
Complete system with FLANN	**7.3731**

Table 3: Results that are obtained by using the same matching threshold τ for each sequence.

Experiments	Precision	Recall
12:10 vs 08:45	0.9942	0.1820
12:10 vs 10:00	0.9974	0.5057
12:10 vs 14:10	0.9976	0.3068
12:10 vs 15:45	0.9946	0.1030

Finally, an interesting way of visualizing the locations where loop closures are detected is in trajectory maps. Fig. 4 visualizes those maps for precisions are at 99.5%. Green stars on the routes indicate correctly closed loops, while false detections are displayed by red lines with circles. In addition to this, Table 3 lists the results that are obtained by using the same matching threshold τ. We particularly report the results at high precisions since false detections have negative effect for reliable localization and mapping [2, 3]. Results show that our method can detect loops under changing illumination with high accuracy in real-time.

Conclusion

Despite the research efforts dedicated to solve the loop closure detection problem, performing accurate and computationally efficient loop closing in real-world environments remains a challenging problem. In this paper, we propose a method for efficient loop closure detection under changing illumination. According to the experiments that we perform in real-world datasets, the proposed method has proven that it can successfully accomplish to detect loops with high accuracy, and it allows real-time operation for long-life localization and mapping.

In future works, we consider to improve the image matching approach which is limited to the specific area in where the agent can traverse. This might be improved by online indexing of previously seen images. Moreover, using trajectory information may increase the loop closing accuracy. Such improvements can be more useful for long-life localization and topological mapping applications.

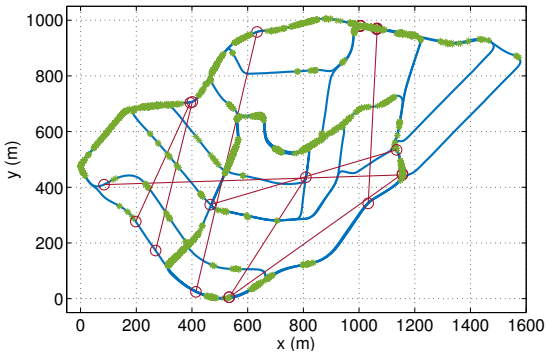
References

[1] Y. Liu and H. Zhang, "Indexing visual features: Real-time loop closure detection using a tree structure," in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3613–3618, 2012.

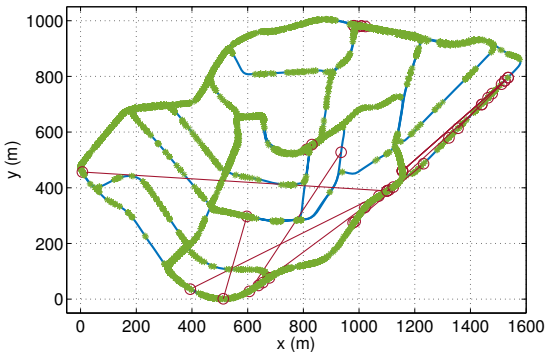
[2] M. Cummins and P. Newman, "FAB-MAP: Probabilistic localization and mapping in the space of appearance," *International Journal of Robotics Research*, vol. 27, pp. 647–665, June 2008.

[3] N. Sünderhauf and P. Protzel, "BRIEF-Gist - closing the loop by simple means," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1234–1241, 2011.

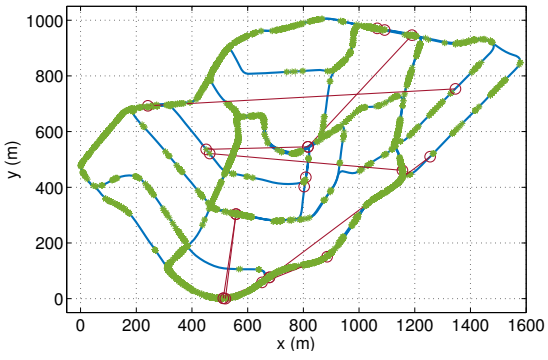
[4] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *IEEE International Conference on Computer Vision (ICCV)*, vol. 2, pp. 1470–1477, 2003.



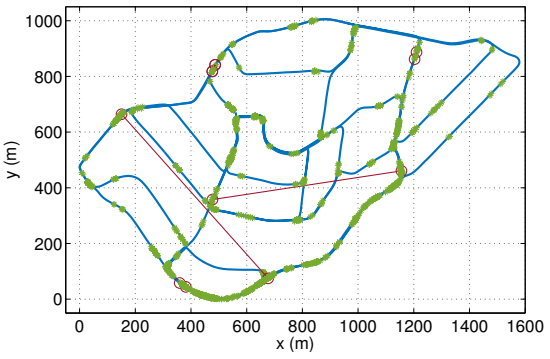
(a) 12:10 vs 08:45



(b) 12:10 vs 10:00



(c) 12:10 vs 14:10



(d) 12:10 vs 15:45

Figure 4: Trajectory maps that are visualized for precision is at 99.5% (see Table 3). Green stars on the routes indicate true positives, while false positives are displayed by red lines with circles.

[5] M. J. Milford and G. F. Wyeth, "SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights," in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1643–1649, 2012.

[6] R. Arroyo, P. F. Alcantarilla, L. M. Bergasa, and E. Romera, "Towards life-long visual localization using an efficient matching of binary sequences from images," *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6328–6335, 2015.

[7] A. J. Glover, W. P. Maddern, M. J. Milford, and G. F. Wyeth, "FAB-MAP + RatSLAM: Appearance-based SLAM for multiple times of day," in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3507–3512, 2010.

[8] W. Maddern, A. Stewart, C. McManus, B. Upcroft, W. Churchill, and P. Newman, "Illumination invariant imaging: Applications in robust vision-based localisation, mapping and classification for autonomous vehicles," in *Visual Place Recognition in Changing Environments Workshop, IEEE International Conference on Robotics and Automation (ICRA)*, 2014.

[9] C. Erhan, E. Sariyanidi, O. Sencan, and H. Temeltas, "An online visual loop closure detection method for indoor robotic navigation," in *IS&T/SPIE Electronic Imaging*, p. 940607, Feb. 2015.

[10] C. Erhan, E. Sariyanidi, O. Sencan, and H. Temeltas, "Patterns of approximated localised moments for visual loop closure detection," *IET Computer Vision*, Dec. 2016 (available online).

[11] Y. Liu and H. Zhang, "Performance evaluation of whole-image descriptors in visual loop closure detection," in *IEEE International Conference on Information and Automation (ICIA)*, pp. 716–722, 2013.

[12] R. Arroyo, P. F. Alcantarilla, L. M. Bergasa, J. J. Yebes, and S. Bronte, "Fast and effective visual place recognition using binary codes and disparity information," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3089–3094, 2014.

[13] E. Sariyanidi, V. Dagli, S. C. Tek, B. Tunc, and M. Gokmen, "Local Zernike moments: A new representation for face recognition," in *IEEE International Conference on Image Processing (ICIP)*, pp. 585–588, 2012.

[14] E. Sariyanidi, H. Gunes, M. Gokmen, and A. Cavallaro, "Local Zernike moment representation for facial affect recognition," in *British Machine Vision Conference*, pp. 108.1–108.11, 2013.

[15] C. McManus, W. Churchill, W. Maddern, A. D. Stewart, and P. Newman, "Shady dealings: Robust, long-term visual localisation using illumination invariance," in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 901–906, 2014.

[16] M. Muja and D. G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," in *International Conference on Computer Vision Theory and Applications*, pp. 331–340, 2009.

[17] M. Muja and D. G. Lowe, "Scalable nearest neighbor algorithms for high dimensional data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 11, pp. 2227–2240, 2014.

Author Biography

Can Erhan received his BS in physics engineering (2012) and his MS in mechatronics engineering (2016) from Istanbul Technical University. He is a researcher in Robotics Laboratory in Controls and Automation Department in the same university. He also works as a software engineer in an R&D company. His interests are computer vision and pattern recognition. His work is focused on long-life localization.