Journal of Theoretical and Applied Information Technology
30th April 2021. Vol.99. No 8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

FIXED PARTITIONING USING AN HEURISTIC APPROACH FOR LOOP CLOSURE DETECTION

AZIZI ABDULLAH[1], IZWAN AZMI[2], MOHAMMAD FAIDZUL NASRUDIN[3], MOHAMMED SALAMEH[4]

[1,2,3] Center For Artificial Intelligence Technology, Universiti Kebangsaan Malaysia, Malaysia

[4] Qatar Science and Technology Secondary School, Qatar

E-mail: [1]azizia@ukm.edu.my, [2]p73064@siswa.ukm.edu.my, [3]mfn@ukm.edu.my, [4]m.salameh1301@education.qa

ABSTRACT

Loop closure detection using visual information needs proper representation to interpret objects in a scene effectively. The scene may contain several objects corresponding to known and new landmarks. The most widely used methods to detect and describe the regions is to use a keypoint detector to localize objects such as in speeded-up robust features (SURF). Most keypoint-based schemes compute salient points on the object based on the curvature principles of geometrical object surfaces. However, not all the object surfaces can distinctively be described using these rules, such as in scenery images that contain high repeating texture features. In the keypoint detector scheme does not consider the flat texture regions resulting in a few detected points which hinder the holistic visual description of images. Thus, we propose to use a fixed-partitioning scheme that divides the image into several blocks for grouping spatial semantic of significance image features. One possible problem in the proposed approach is to identify the number of partitions and partition size for image description. Thus, an heuristic approach is used to identify these parameters for loop closure detection. A famous computational expensive Real-Time Appearance Based Mapping (RTAB-Map) simultaneous localization and mapping (SLAM) is used to validate the proposed scheme. The results show that the proposed approach outperforms the standard keypoint detector on two datasets, namely Lib6 Indoor and New College

Keywords: *Vision-based SLAM, Loop-closure detection, localization, Fixed Partitioning*

1. INTRODUCTION

Autonomous mobile robots (AMR) are intelligent agents that can perform desired tasks automatically. The tasks can be performed using mechanisms that allow the robots to navigate through a real-world environment. One of the challenging tasks in AMR is localization. The robot's position needs to be determined from the acquired map of an environment. For estimating its location, the robot uses sensor such as laser scanners or visual cameras which enable it to model the real-world environment. However, laser or optical scanners depends on line-of-sight. Besides that, it cannot see undercuts or hidden surfaces that hinder constructing a complete map of an environment. Thus, many robotics researchers are moving to use visual cameras to provide more rich information for describing the environment. Following this, many researchers in machine vision have proposed many useful visual descriptors for dealing with the complex problem of handling high dimensional pixel representations.

This work focuses on loop closure detection for vision-based SLAM using a heuristic algorithm based on fixed partitioning and RTAB-Map. In literature, the loop closure detection algorithms aim to recognize a previously visited place from current visual sensors. These algorithms can be used in robotics, such as to solve the localization and kidnapping problems. The method is proposed mainly due to some limitations in the traditional approaches in understanding, recognizing, and validating surrounding environments. Most standard procedures use probabilistic schemes in strictly Cartesian space to precisely describe a complex 3-D geometrical model of the environment. Using 3D for modeling the surrounding environment is very complicated, especially to recover topology of the ground from noisy real-world objects and sensors. All these, of course, can increase the problems of maintaining a

global map for localization. Therefore, the need for more robust for environment description and preserve only its topology became unavoidable.

Choosing and using digital cameras for choosing and using digital cameras for capturing visual information for SLAM have now become widespread. This type of sensors is cheap and easy to configure and easy to apply computer vision algorithms to do complex vision tasks for robotics. One possible task is to use the algorithm in autonomous mobile robots to describe a scene of different views and orientations for visual SLAM (VSLAM).

In VSLAM, one of the crucial parts to enhance the mapping of VSLAM algorithms is to use a loop closure detection approach. This problem consists of a series of tasks to detect previously visited places in an environment from current visual sensors. Thus, once a loop closure is detected, the actual robot pose can be precisely estimated. It is useful for solving global localization and kidnapping in a deterministic environment. As a result, it attracts many computer vision researchers to research loop closure detection using vision sensors. One of the earlier works on using visual SLAM for mobile robot localization is from Ulrich and Nourbakhsh [1].

In [1], they have introduced the concept of appearance-based representation for visual place description and a similarity measure such as L1 or L2 distance to obtain similarities of the location choices. The concept that they used is similar to content-based image retrieval systems, which are to retrieve past information or images from a database. The images are represented and indexed by their visual content such as color, texture, and shape in the system. After that, a similarity measure is used for matching existing world images in the database. Besides using these features, it is also popular to use the local appearance approach by clustering feature vectors extracted from local points features into similar group patterns. Following this, Newman et al. [2] have proposed to use the visual appearance and laser ranging sensor for outdoor SLAM.

In [2], a sequence of images from a camera is used to detect loop closure using a novel appearance-based detection process. They reported that the method is robust to repetitive visual structure and provides a probabilistic measure of confidence. In [3] proposed to use the most common local appearance-based approached, namely bag-of-features for loop closure detection. In this work, they found that the visual location can be easily identified using a set of unordered cluster features. In this work, they found that the visual location can be easily identified using a set of unordered cluster features. The features are computed from a local image features to represent patches such as SIFT [4] or SURF [5] for vector descriptor and ORB [6] for binary descriptor. A clustering technique is then applied to similar group patterns into a cluster visual codebook that explicitly considers object category information. In the loop closure detection, the codebook can be constructed offline using identified location images for training or online using an incremental clustering algorithm [7]. Besides using similarity measures for landmark matching, in [8] [9] proposed to use the Bayesian filter to estimate the loop closure from a previously visited place. Similar to previous methods, visual words are used to the symbolic representation of visual places.

One of the most critical problems in a robot navigation system for mapping is identifying the sub-images representing landmarks. However, this operation looks relatively easy for people but surprisingly difficult for robots. Therefore, to exploit the actual benefit of loop closure detection, we proposed the use of a popular and computationally expensive system named as the Real-Time Appearance Based Mapping (RTAB-Map) [9]. RTAB-Map can be classified as a type of RGB-D SLAM approach for loop closure detection having real-time computational constraints [10]. Thus, based on the literature survey of some technical aspects of the current VSLAM/SLAM, The need of robust based on a feature description of content is key to improve the loop closure detection performance. This technical aspect is important in such as querying, feature matching, indexing and storing of landmark images.

In general, most of the works in VSLAM rely on two image schemes namely salient points detection (local descriptor) and the global image descriptor. However, one problem in the salient points image scheme is that it rely a lot on the foreground information. Thus, if the image landmarks contain more background information such as in the scene environment, it may hinder the scheme works optimally. But, in the global image scheme, the background information is more informative to be used for image description than the foreground. Thus, we believe the propose fixed-partitioning has an ability to balance between these two popular schemes and has semantic significance in image representation. In this work, we propose to improve

the loop closure detection performance by reducing RTAB-Map computational complexity with the fixed partitioning scheme.

In many RGB-D slam approaches, the salient points detector scheme has shown an excellent performance in improving the landmarks. However, one problem in the scheme is that most of the detected points are came from the high distinctive geometrical structure that corresponds to the edge, corner, or texture transition of the image. Nevertheless, not all images contain such components for thoroughly describing landmarks. It would result in coarse indexing when the few local point samples were detected [11]. Therefore, in this paper, we propose to use a fixed partitioning scheme that divided landmark images into sub-regions and extracts local image features from each region for description. After that, the bag-of-words descriptor is used to describe the landmark images via an online clustering algorithm. Using the fixed partitioning scheme would give an accurate interpretation of an important region that correspond to objects or part of a landmark and would give better image description [12] [11].

The contribution of our work is: (1) we present the fixed partitioning scheme for dividing an image into sub-images for efficient landmark description for loop closure detection using a heuristic based on RTAB-Map. (2) We demonstrate the effectiveness of the bag-of-words approach from the fixed partitioning scheme. (3) We compare the proposed method with the salient point's detection scheme on three popular loop closure datasets.

The rest of the paper is organized as follows. Section 2 describes the related work of the proposed approach. Section 3 describes our system for real-time appearance-based mapping. Experimental results on the Lib6indoor, city center, and new college datasets are shown in section 4. Section 5 discusses the results and concludes the paper.

2. RELATED WORK

Recently, loop-closure detection methods using point cloud descriptors have been introduced in many studies and have been shown to outperform on many public indoor and outdoor navigation datasets [13] [14] [15].The loop closure detection algorithms aim to recognize a previously visited place from current visual sensors. These algorithms can be used in robotics, such as to solve the localization and kidnapping problems. In literature, most SLAM algorithms work with sensory data such as laser to capture surrounding information or odometry, which provides the robot's wheels' rotational speeds for building a map. The map is then used to estimate the robot's location by comparing the current view of the environment's previously recorded images.

2.1 Visual Simulatenous Localization and Mapping

SLAM is an approach for a mobile robot to construct a map of the surrounding environment. At the same time, it recognizes the robot's location on the map [16]. The use of robotic visual sensors in SLAM is essential in perceiving and obtaining the environment's features. They also serve as a crucial source for constructing a reliable map for autonomous navigation in a new indoor or outdoor environment [17]. The visual sensors have different types that produce various features of the surrounding environment. For an efficient VSLAM approach, a type of sensors must be carefully chosen according to the environment's nature to produce reliable visual information on which the mapping method depends.

Generally, VSLAM/SLAM contains two essential processes, namely (a) Mapping - is a registration process for observing landmark features captured by a robot from the surrounding environment and saving them in a structural manner that mimics the environment's landmark positions to each other, and/or according to a reference position in correlation with other landmarks to be used for a robot's diverse functions, and (b) Localization - is the process of using the observed sensor readings of a robot to identify the location of the robot on the map which is instantly built on time according to the movement of the robot [18] [19]. The map building depends on the sensor data association in detecting whether the observed features belong to a new location or the location already saved on the map. This process is known as Loop Closure Detection (LCD), which detects the robot's position and reduces the location uncertainty.

2.2 Loop Closure Detection (LCD)

The correct perception of the surrounding environment is one of the essential aspects of VSLAM. Although different visual sensors can carry out this task, all of them have some problems with noise. The sensor noise directly affects mapping and localization efficiency, as VSLAM relies only on raw sensor data. Therefore, the graphical LCD is highly required for accurate mapping and localization. LCD is trying to find a

match between the current and previously visited locations on a SLAM map. As a result, the robot will be able to reduce uncertainty locations. It will construct a consistent representation of the environment [20] [9].

2.3 Local Image Features

In computer vision, local features are different in small areas of a region. One of the famous principles to measure a characteristic of local features is the principle of curvature. It measures the selected geometric attributes of one object from another in object recognition. The geometric properties can contain a specific shape of a small boundary segment or a surface patch. The most popular local features arguably SIFT and SURF descriptors. SIFT descriptor [4] constructs the histograms of gradient orientations computed around the points as the descriptor. SIFT uses an interest point detector to detect salient locations that have specific repeatable properties. After that, a 128-bin was used for image description. Another local descriptor is SURF [5], which computes the sum of the Haar wavelet response around the image description point of interest. In contrast to SIFT, it uses the integral image for approximating the second-order derivatives for points detection and generates a 64-bin for image description.

3. THE PROPOSED METHOD

3.1 Real-Time Apperance-Based Mapping (RAB-MAP)

RTAB-Map is a memory management approach for the real-time appearance-based loop closure detection [13] [21].It uses the local appearance descriptor for description and the discrete Bayesian filter for evaluating the loop closure hypothesis. In this approach, the map is constructed by linking new acquire images with previous ones based on the loop closure probability values. This process is fully incremental; i.e., it starts from an empty model and learns from the first training visual image to construct a complete map.

RTAB-Map uses (a) Working Memory (WM) to keep the most recent and frequent observed locations, (b) Long Term Memory (LTM), to store other observed locations. When a match is found between the current location and the one stored in WM, an associated location can be remembered and updated. Then (c) Short Term Memory (STM) stores the poses and retrieves them from the long term memory when the loop closing is required. Fig. 1 shows RTAB-Map memory management and loop closure detection steps [22].

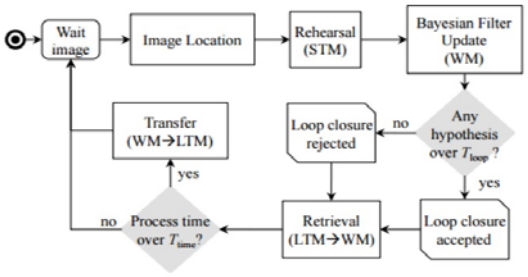

Figure 1: Architecture of the RTAB-Map memory management loop closure detection steps.

RTAB-Map uses the bag of words (BoW) to make a visual dictionary of local features. The BoW is constructed using an online incremental algorithm with SURF descriptors in a KD-tree structure. The codebook is built and updated using the Fast Library for Approximate Nearest Neighbors (FLANN) with Nearest Neighbor Distance Ratio (NNDR) [23]. After that, this feature descriptor is used to recognize locations. It uses a topological map with nodes representing the visited locations. Each node contains the location signature, a set of visual features extracted from the image belonging to a location.

The location is detected through four main stages, starting with the Sensory Memory (SM) are extracted from the current image and generate signature using online BoW to create a new node. The second stage is Short Term Memory (STM) with the First-In-First-Out (FIFO) structure and a fixed length. Its primary duty is to merge two neighboring locations similarity matching. When the STM stack is full, the stack's first saved location will pass to the next stage. The third stage is Working Memory (WM), which is the active part of the memory where a location is identified as a new location or as a loop closure to a previous location. RTAB-Map uses the Bayesian filter to estimate the full posterior probability. In the final stage, if the current location is registered as a loop closure, WM will retrieve the two neighbors for the current loop closure location. These neighbors have a high probability of being the next loop closure.

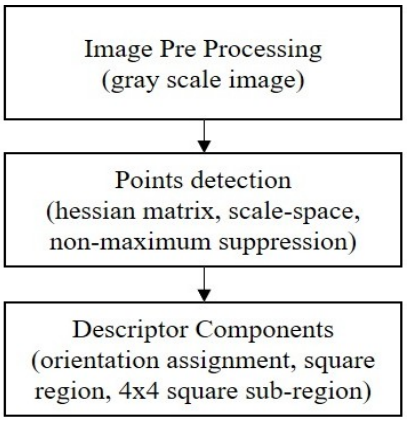

Figure 2: SURF's process flow for interest point detection and description.

3.2 Speed Up Robust Features (SURF)

SURF is known as an algorithm capable of finding correspondence between images but in a fast way. The SURF algorithm constructs an orientation and scale-invariant detector and descriptor for a given image. The overall process is illustrated in Fig. 2.

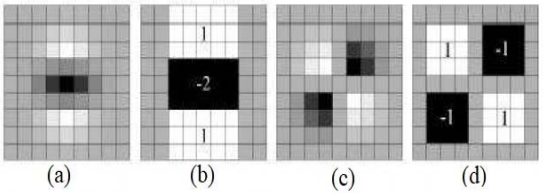

Figure 3: (a) Left: Gaussian second order partial derivatives in y-direction. Right: approximation of using box filter. (b) Left: Gaussian second order partial derivatives in xy-direction. Right: approximation of using box filter

Image Preprocessing - Usually, interest points are detected under illumination changes in an image. Therefore, the first step is to convert color images to grayscale images. The grayscale type is used because it is simple to interpret and enhance. Besides using the grayscale values for every pixel, each image in the dataset is also resized to increase the performance of the points detector algorithm. Therefore, particular input images are down-sampled to decrease the number of pixels, while maintaining its aspect ratio so that the image quality can almost be preserved.

Points Detection -Once the image is transformed into the grayscale level, the next step is to localize the interest points. The SURF point detector is based on the Hessian matrix. Given a point x= (x, y) in an image I, the Hessian matrix $H(x,\sigma)$ in x at scale σ is defined in (1) as follows

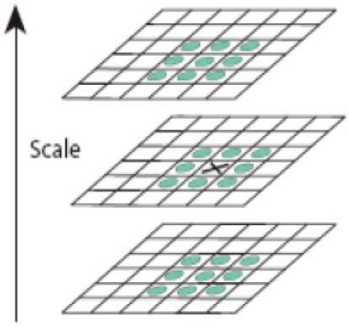

Figure 4: Non-maximum suppression: An interest point is compared to its 26 neighbours.

$$H(x,\sigma) = \begin{bmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{yx}(x,\sigma) & L_{yy}(x,\sigma) \end{bmatrix} \quad (1)$$

where $L_{xx}(x,\sigma)$ is the convolution of the Gaussian second order derivative $\frac{\partial^2}{\partial x^2} g(\sigma)$ with the image I in point x, and similarly for $L_{xy}(x,\sigma)$ and $L_{yy}(x,\sigma)$. SURF uses the regular Gaussian convolutions to approximate Laplacian of Gaussian (LoG) with simple box filters. Thus, to filter higher layers, the filter sizes are successively increased. This algorithm is done without down sampling for higher levels, resulting in images of the same resolution used on each scale. For example, we use the filter box of size 9×9 to approximate to Gaussian derivative with scale or σ=1.2, and the 27×27 filter is equivalent to Gaussian derivatives with σ=3.6. The use of box filters to convolve the original image at different scales is possible due to the use of integral images [24] that allow the computation of rectangular box filters in near-constant time. Figure 3 (a) and Figure 3 (b) show corresponding Gaussian second- order derivatives with the box filters in the y-direction and xy-direction, respectively.

Once the approximation of second-order Gaussian derivatives is determined, the next step is to use a non-maximum suppression in a 3×3 neighbourhood as indicated in Figure 4 to identify key points. In this step, each sample point is compared with eight neighbours on the same scale, nine neighbours in the above and nine neighbours in the below as shown in Figure 4. In short, 26 points have to be compared at a time. A point is

selected as a salient point if it has the largest or the smallest value. Accepting or rejecting the location and scale of interest points is relying on the determinant of the Hessian. Let us denote the approximation of the second-order derivatives as D_{xx}, D_{yy} and D_{xy} which are computed by applying the different simple box-filters. Next, the box filters' weights are chosen adequately to approximate the Hessian's determinant as in (2) as follow:

$$Det(H_{approx}) = D_{xx}D_{yy} - (0.9D_{xy})^2 \quad (2)$$

After that, the ratio of principal curvature value is measured and compared to below some threshold. Finally, the found maxima of the determinant of the approximated Hessian matrix are interpolated in scale and image space. For more details, see [5].

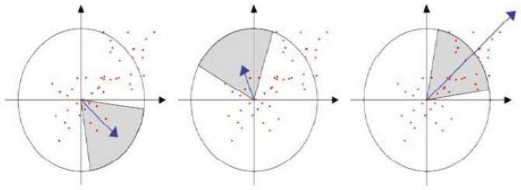

Figure 5: SURF's dominant orientation assignment. The longest vector shows the dominant orientation.

Point Description - The SURF descriptor is constructed in two steps. The first step consists of fixing an orientation based on information from a circular region around the detected key point. After that, a square area is constructed and aligned to the selected orientation and extract the SURF descriptor from it. To be invariant to rotation, SURF tries to identify an orientation for the interest points. The orientation is determined through two steps:

(a) SURF first calculates the Haar-wavelet responses in *x* and *y* directions in a circular neighborhood of radius 6s around the key point. The s is the scale in which the key point was detected.

(b) After that, the sum of vertical and horizontal wavelet responses is calculated in a scanning area, then changes the scanning orientation by adding $(\frac{\phi}{s})$ and re-calculates until the orientation has the highest sum value. This orientation is the main orientation of the feature descriptor is determined. Figure 5 shows the orientation is computed and selected. The following step is implemented:

(i) The first step consists of constructing a square region centered around the identified keypoint and its orientation.

(ii) The region is split up regularly into smaller 4×4 square sub-regions. Furthermore, for each sub-region, a few simple features at 5×5 regularly spaced sample points are computed. In this case dx the Haar wavelet response in the horizontal direction and dy the Haar wavelet response in the vertical direction. To increase the robustness towards geometric deformations and localization errors, the responses dx and dy are first weighted with a Gaussian (σ = 3.3s) centred at the key point. Figure 6 shows how the descriptor is computed in SURF.

Then, the wavelet responses *dx* and *dy* are summed up over each sub region and form a set of entries to the feature vector. To bring in information about the polarity of the intensity changes, the sum of the absolute values of the responses, |*dx*| and |*dy*| are extracted. Hence, each sub-region has a four dimensional descriptor vector v for its underlying intensity structure V = (∑dx, ∑dy, ∑|dx|,∑|dy|). This results in a descriptor vector for all 4 x 4 sub-regions of length 64-bin.

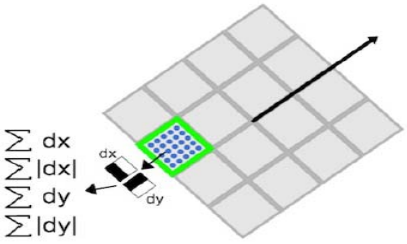

Figure 6: SURF's description with the dominant orientation

3.3 Bayesian Learning

A discrete Bayesian filter is used in loop closure detection to estimate the similarities between the current location and the locations that have been visited. A decision is then taken to register the current location as a new location (never visited) or as a loop closure to a previous location (already visited). The Bayesian filter tries to estimate the full posterior probability $p(S_t|L^t)$ for the current location L_t at time t, where S_t is a set of all the loop closure candidates for the location L_t. If the

locations $L_t$and $L_i$are representing the same location, where $i \in [0, \ldots, n]$ and n is the number of locations in the map, then the probability of the loop closure between L_t and L_i will be S_t= i. On the other hand, if the location L_t is a new location, then S_t= -1. The full posterior probability for all $i = -1, \ldots, t_n$ can be termed as (3).

$$(S_t|L^t) = \eta \underbrace{p(L_t|S_t)}_{Observation} \underbrace{\sum_{i=-1}^{t_n} \underbrace{p(S_t|S_{t-1}=i)}_{(Transition)} p(S_{t-1}=i|L^{t-1})}_{Belief} \quad (3)$$

where η is a normalization term and L^t is for locations in the RTAB-Map working memory (the search space of the Bayesian filter), then $L^t = \{L_{-1}, \ldots, L_t\}$ and tn is time index. For more information we refer to this paper in [13].

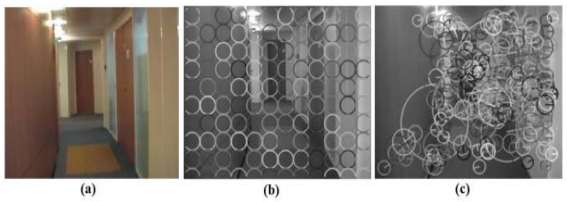

Figure 7: (a) original image (b) the proposed fixed-partitioned image scheme and (c) the popular interest points scheme.

3.4 Heuristic Approach for Fixed Partitioning Scheme

In many images, the grey values are not generally constant across the whole image. Thus, the images are often divided into parts before features are computed from each part. This idea is based on the assumption that the image can be composed into partitions. A simple function can model in each partition. There are many ways to partition the image. One possible approach is to divide the image in tiles of equal size and summarize each tile's dominant feature values. After that, the SURF descriptor of size 64-bin is used to describe visual information for each region. Moreover, we believe that the combination of all partition regions of the partitions can compensate for the complete loss of spatial information of BoW. However, this rule is often violated by principle curvatures to determine local features. However, this violation in itself has semantic significance [12]. The scheme has been explored in many computer vision applications such as in [25], [26], [27] and shown remarkable performance in describing image content. In this work, the fixed partitioning scheme is proposed to divide equal size images, as shown in Fig. 7 (b). One possible problem in this approach is that to identify the number of partitions and partion size. Thus, a simple heuristic approach is used to determine the optimal values of these important parameters. In this approach, we have tested it using a set of different settings of partitions and sizes. The partition size depends upon the size of the image (Fig 7(b)). The best setting then will be used to employ the fixed partitioning in the RTAB-Map. We used the fixed partitionig scheme with the heuristic approach due the following reasons:

(a) It is simple and needs less overhead of implementation and computation. Thus, it may reduce some computational expenses on online SURF for extracting local features from images. Also, some VSLAM tasks such as vision processing and map building required very high computational power to achieve real-time performance. Besides, RTAB-Map has some real-time constraints, especially on the memory management architecture that increases its overall efficiency [10].

(b) Different fixed partitioning schemes such as 4 x 4, 8 x 8, etc. can be tested to capture the input image's informative description. After partitioning, low-level visual features are computed for each region. These features are quantized using an online clustering algorithm. Often, the number of partitions is less than the number of patches detected using the salient point detector (SURF). Thus it gives some advantages to improve quantization performance for online BoW.

(c) Besides the number of partitions, the partitioning scheme's size is much smaller than the actual image size. Thus, it gives less computational time to visit all partitions for calculating spatial information in the image. However, problems might arise if the fixed partitioning divides an important region into two or more parts. Thus it might give some advantages to the saliency-based approach. The proposed fixed partitioning scheme is proposed in Fig. 8.

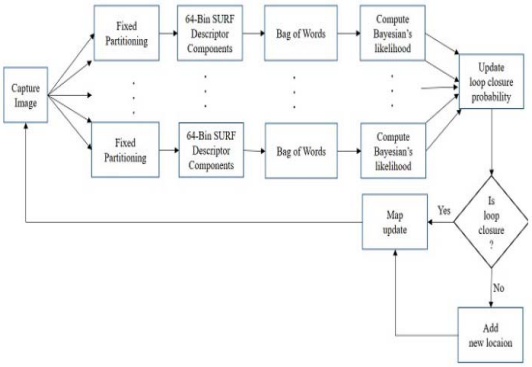

Figure 8: Fixed partitioned scheme for RTBA-Map

3.5 SURF Point Detector and Descriptor

Images taken from scenes and objects usually have many variabilities such as viewpoint, clutter, and occlusion. Most of these problems are quite challenging to handle with a global-based approach like segmentation or fixed partitioning. There exists a technique that can cope with these problems named the saliency-based approach. The approach is claimed to be local, and so it is robust to occlusion and clutter. Besides that, it is also invariant to image transformations and illumination changes. Furthermore, the algorithm does not need prior segmentation of the images. However, it is based on the repeatable computation of local extrema points between an image's scale-spaces. This approach's main idea is to find the most informative locations of salient points in an image. There are several algorithms to achieve this goal such as scale invariant feature transforms (SIFT) [4] and speeded up robust features (SURF) [5]. The SURF algorithm describes the salient points by dividing images into various informative regions or patches. The patches processed recursively are composed of different sizes and locations, as shown in Fig. 7 (c). A histogram of size 64-bin is used to describe visual information for each region.

4. EXPERIMENTAL RESULTS AND ANALYSIS

In this paper, we have implemented the fixed partitioning approach with SURF descriptor for loop closure detection. We have tried several partition for example 4 x 4, 10 x 10, 20 x 20, etc. We found that the partition of size 20 x 20 gives the best performance and is used to index all the circular regions (Fig. 7(b)). After that, a histogram of size 64-bin is used to describe each region. For salient point detector (Fig. 7(c)), SURF with the default settings are used for all experiments. Similar to the fixed partitioning approach, the histogram of 64-bin used to describe each patch. The proposed approach is implemented on an Intel(R) Xeon(R) CPU @ 2.83GHz machine of 12GB RAM in the Debian GNU/Linux 7 platform. We used OpenCV and RTAB-Map libraries for processing images and VSLAM, respectively. The C++ programming language is used for the entire algorithm development.

4.1 Dataset

Public datasets are used to evaluate the proposed algorithms' capabilities and robustness under various conditions, including indoor and outdoor environments. The three public datasets: Lip6 Indoor, City Centre, and New College have been utilized in this research. These datasets are widely used to evaluate the VSLAM algorithms, which are considered challenges since they involve a highly similar image, features in the indoor environment, and the outdoor environment contain fluctuating scenes. Another reason for choosing these datasets is the frame rate of the image capture, which is compatible with the proposed algorithms.

(a) Lip6 Indoor: Lip6 Indoor is the first part of the dataset that is collected by [7]. Lip6 Indoor dataset contains 388 images of two loops in medium-sized corridors with corners. The image size of240 192 which is captured at 1 Hz using a single-monocular hand-held camera with a 60-degree field of view. The illumination is constant under artificial lighting conditions, Fig. 9 shows sample images from this dataset. Some of the dataset challenges are that they contain high perceptual aliasing and perspective transformation, such as the changes in scale and the point of view. Fig. 9 show some images from the dataset.

Figure 9: Samples of lip6 Indoor dataset

(b) City Center: City Centre dataset is a subset of the Oxford dataset. The City Centre dataset

contains 2474 images with a size of 640 480 acquired from two cameras (left and right). The images are captured outdoors along 2 km in public roads at a rate of 0.5 Hz. The images include dynamic objects which are taken on a windy day in bright sunshine that show the foliage and shadows in dynamic motion. To evaluate the performance of the proposed LCD algorithms' performance, which requires a single image, the two images, left and right, have been concatenated into a single image of a new size of 1280 480. The total number of sequential images is 1237 of two rounds. The authors provide the ground truth map of the loop closure locations, which was manually signed by themselves. Fig. 10 shows samples of images taken from the dataset.

Figure 10: Samples of City Centre dataset

(c) New College: The New College dataset [28] includes images from a platform driving around the campus. The dataset consists of 8127 frames, which have been collected over a 2.2km long trajectory. However, no real ground truth is avail-able for this dataset. The data was collected using a wheeled robot, a Segway RMP200. The data is available at the dataset website http://www.robots.ox.ac.uk/NewCollegeData. Fig. 11 shows some images from New College dataset.

4.2 Performance Measurement and Discussion

The proposed loop closure detection algorithm was evaluated using precision. Precision is defined as the ratio between the number of correctly detected loop closure frames (True- Positive (TP)) and the total number of detected loop closure frames (True-Positive and False-Positive (FP)) as in eq (4) as follows:

Figure 11: Samples of New College dataset

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

Table 1 shows the precision results of RTAB-Map with salient points and fixed partitioning schemes on different datasets. According to the table, it is clearly seen that the precision of fixed partitioning outperforms salient points on two datasets, namely lib6 indoor and New College datasets.

TABLE I: The precision of fixed partitioning and salient points classification results. In each row, the best result is bolded. L6I: Lip6 Indoor, NC: New College, CiC: City Center. The best result is bolded.

Methods	L6I(%)	NC	CiC
Salient Points	23.47	59.17	**95.75**
Fixed Partitioning	**24.25**	**69.07**	91.85

Both approaches perform worse on the lib6 indoor possible due to confusing of many landmarks as indicated in Fig 9. In this case, the lack of spatial coherence between partitions or regions in both methods is possibly the main reason. However, the fixed partitioning gives slightly lower results on City Center dataset. The lower result is perhaps due to the City Centre dataset more explicitly and relies less on background information. Besides, there are two factors that influence the fixed partitioning scheme namely the number of partitions and image descriptor. These factor will indirectly affect the recognition of landmark performance.

5. CONCLUSION AND FUTURE WORK

In this paper, a fixed partitioning approach for loop closure detection is presented. The scheme divides an image into several blocks of a fixed size. After that, the SURF descriptor of size 64-bin is used to describe each partition for landmarks

identification. One problem in the approach is to identify the number of partitions and partition size. Thus, a simple heuristic approach is used to determine these optimal parameters.

The experiments shown that partition size of 20 x 20 gave the best performance for loop closure detection and the partition size depends upon the size of the image. By using this setting, it shows that the fixed partitioning can extract a semantic significance which is useful for robot visual perception tasks. Besides, it can cover the entire image area for holistic image description. The proposed approach is evaluated on different publicly available outdoor and indoor datasets.

The results show that the proposed method is capable of identifying landmarks that have been visited. In future, we want to add more features and combine with different partition sizes for landmark description. The proposed method shown that the fixed partitioning outperforms on the datasets namely L6I and NC, but performs a bit worse on the more difficult CiC dataset. Therefore, it would be interesting to model landmarks more explicitly in this dataset. Besides, our fixed partitioning approach calls upon a Bayesian filtering framework with likelihood computation for making decision and should be extended to the real implementation of VSAM algorithm.

ACKNOWLEDGEMENT
This work has been supported by the Malaysia's Ministry of Higher Education Fundamental Research Grant
FRGS/1/2019/ICT02/UKM/02/8 and
ERGS/1/2011/STG/UKM/02/41.

REFRENCES:

[1] I. Ulrich and I. Nourbakhsh, "Appearance-based place recognition for topological localization," Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No.00CH37065), vol. 2, 2000, pp. 1023–1029 vol.2.

[2] P. Newman, D. M. Cole, and K. Ho, "Outdoor slam using visual appearance and laser ranging," Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006, 2006, pp. 1180– 1187.

[3] D. Nister´ and H. Stewenius,´ "Scalable recognition with a vocabulary tree," 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), vol. 2, 2006, pp. 2161–2168.

[4] G. Lowe David, "Distinctive image features from scale-invariant key-points," International Journal of Computer Vision, 2004.

[5] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-up robust features (surf)," Comput. Vis. Image Underst., vol. 110, 2008, pp. 346–359.

[6] E. Rublee, V. Rabaud, K. Konolige, and G. R. Bradski, "Orb: An efficient alternative to sift or surf," 2011 International Conference on Computer Vision, 2011, pp. 2564–2571.

[7] A. Angeli, D. Filliat, S. Doncieux, and J. Meyer, "Fast and incremental method for loop-closure detection using bags of visual words," IEEE Transactions on Robotics, vol. 24, 2008, pp. 1027–1037.

[8] A. Angeli, S. Doncieux, J.-A. Meyer, and D. Filliat, "Real-time visual loop-closure detection," 2008 IEEE International Conference on Robotics and Automation, 2008, pp. 1842–1847.

[9] M. Labbe,´ "Rtab-map as an open-source lidar and visual slam library for large-scale and long-term online operation," 2018.

[10] R. Doriya, "Development of a cloud-based rtab-map service for robots," 2017 IEEE International Conference on Real-time Computing and Robotics (RCAR), 2017, pp. 598–605.

[11] A. Abdullah, R. C. Veltkamp, and M. A. Wiering, "Fixed partitioning and salient points with mpeg-7 cluster correlograms for image categorization," Pattern Recognitionion., vol. 43, 2010, pp. 650–662.

[12] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, 2000, pp. 1349–1380.

[13] M. Labbe and F. Michaud, "Appearance-based loop closure detection for online large-scale and long-term operation," IEEE Transactions on Robotics, vol. 29, 2013, pp. 734–745.

[14] M. A. Uy and G. H. Lee, "PointNetVLAD: Deep point cloud based retrieval for large-scale place recognition," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 4470–4479.

[15] M. O. Salameh, A. Abdullah, and S. Sahran, "Ensemble of Bayesian filter with active and passive nodes for loop closure detection," 2017 18th International Conference on Advanced Robotics (ICAR), 2017, pp. 482– 486.
[16] J. Fuentes-Pacheco, J. R. Ascencio, and J. M. Rendon´-Mancha, "Visual simultaneous localization and mapping: a survey," Artificial Intelligence Review, vol. 43, 2012, pp. 55–81.
[17] J. Aulinas, Y. R. Petillot, J. Salvi, and X. Llado,´ "The slam problem: a survey," in CCIA, 2008.
[18] E. Garcia-Fidalgo and A. Ortiz, "Vision-based topological mapping and localization methods: A survey," Robotics Auton. Syst., vol. 2015, 64, pp. 1– 20.
[19] T. Taketomi, H. Uchiyama, and S. Ikeda, "Visual slam algorithms: a survey from 2010 to 2016," IPSJ Transactions on Computer Vision and Applications, vol. 9, 2017, pp. 1–11.
[20] A. Angeli, S. Doncieux, J.-A. Meyer, and D. Filliat, "Incremental vision-based topological slam," 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2008, pp. 1031–1036.
[21] M. Labbe´ and F. Michaud, "Online global loop closure detection for large-scale multi-session graph-based slam," 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2014, pp. 2661–2666.
[22] ——, "Memory management for real-time appearance-based loop closure detection," 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2011, pp. 1271–1276.
[23] M. Muja and D. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," in VISAPP, 2009.
[24] P. Viola and M. J. Jones, "Rapid object detection using a boosted cascade of simple features," Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, vol. 1, 2001, pp. 511–518.
[25] A. Abdullah and M. A. Wiering, "Cirec : Cluster correlogram image retrieval and categorization using mpeg-7 descriptors," 2007 IEEE Symposium on Computational Intelligence in Image and Signal Processing, 2007, pp. 431–437.
[26] I. K. Sethi, I. Coman, B. Day, F. Jiang, D. Li, J. L. Segovia-Juarez, G. Wei, and B. You, "Color-wise: a system for image similarity retrieval using color," in Electronic Imaging, 1997.
[27] K. V. D. Sande, T. Gevers, and C. G. M. Snoek, "Evaluating color descriptors for object and scene recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, 2010, pp. 1582–1596.
[28] M. Smith, I. C. Baldwin, W. Churchill, R. Paul, and P. Newman, "The new college vision and laser data set," The International Journal of Robotics Research, vol. 28, 2009, pp. 595 – 599.