HS Combined Histogram for Visual Memory Building and Scene Recognition in Outdoor Environments

K. Rebai, O. Azouaoui and N. Achour

***Abstract*— A novel appearance-based approach to perform scene recognition for loop closure detection is proposed in this paper. The proposed solution contains two modules: the image descriptor extraction and a visual memory for scene recognition. This approach employs a combined HS histogram as a global image descriptor. The visual memory composed of view cells has the architecture of Fuzzy ART neural network. The use of this kind of neural network allows fast, incremental, competitive and unsupervised learning of a visual space representation. The effectiveness of our framework has been proven on a challenging large-scale suburban datasets.**

I. Introduction

Nowadays scene recognition for mobile robot place recognition and loop closure detection has become one of the most active research areas [12][18][29][33]. It plays an important role for robot localization [2][33][42] and front-end visual SLAM problems [18][23][31][39][43]. However, for real and long life operation, the mobile robot shall be able to deal with environmental appearance changes. Therefore, it is much more difficult to extract compact and robust image descriptor for urban environments, and to compute similarity between them.

For a scene recognition system, the extraction of visual information for image description and the search of similar descriptor are extremely important. Many efforts have been made over the last decades to find a suitable image descriptor in order to obtain an effective image descriptor in terms of robustness and complexity. The developed descriptors are typically separated in global or local descriptor. Each descriptor has its advantages and its drawbacks, which motivates again researches in computer vision for image retrieval and robotics problems to continue their investigations in this area.

The purpose of this paper is to improve appearance-based SLAM in long-term operation and in dynamic outdoor environments. To deal with the problem of vision-based loop closure detection, our solution is based on the primate's visual memory features. The proposed image descriptor consists on computing global descriptor within local regions in the HSV color space. This descriptor is the input of a computational model of the visual memory. The design of the bio-inspired visual memory takes into account some properti<es of the human memory. The proposed approach does not require any a priori knowledge of the scene and does not need anyoffline-learning phase.

The remainder of this paper is organized as follows: Section 2 reviews relevant related works. Section 3 gives an overview of the proposed computational model of the visual memory. Then, Section 4 introduces a simple and effective global image descriptor using HS combined histogram. Section 5 shows the experimental results and discusses the performance of the proposed approach in comparison to other methods. Furthermore, Section 6 presents the main conclusions of this paper and.

II. Related Works

Ulrich and Nourbakhsh [45] have introduced the concept of appearance-based place recognition and discussed the similarity between place recognition and image retrieval problems. Therefore, appearance-based approaches for place recognition and loop closure detection have gained increasing attention in the last years. These approaches can be broadly classified according to the used image descriptor. The first class use local descriptors while the second employs global image descriptors.

Local descriptor are computed around particular points in the image called keypoints. The most common local descriptors are SIFT [24], SURF, BRIEF [5], ORB (Orientation BRIEF) [41], and BRISK [21]. Bag of word is the most popular framework that use local descriptors for scene recognition. This concept proposed by Nister et al. [34] was largely exploited for loop closure detection [1][19][20]. It suggests that the images can be represented as a set of unordered elementary features (local descriptors) [1] taken from a visual dictionary. The place recognition consists of two processing phases: offline dictionary building and online similarity test between the extracted words and the available dictionary. To overcome the problem of offline dictionary building, Angeli et al. [1] proposed an incremental and online dictionary building. In [19], a vocabulary tree is proposed to limit the search for similar images without comparing them with all the available images in the database. A similar idea was also discussed in [4] and known as "tree of words" which is a hierarchical approach to bag of words.

The fast appearance-based mapping (FAB-MAP) has been proposed by Cummins and Newman [11] for appearance description . The first version of FAB-MAP [11] allows closed loop detection with small recall (16% to 37%) for 100% precision, while its extensions: *Accelerated* FAB-MAP

K. Rebai was with Centre de Développement des Technologies Avancées, Algiers, Algeria. She is know with Ecole Nationale Supérieure de Technologie and Université des sciences et de la technologie Houari Boumediene, Algiers, Algeria. (e-mail : karima.rebai@gmail.com)

O. Azouaoui is with Centre de Développement des Technologies Avancées, Algiers, Algeria. (e-mail : azouaoui@rocketmail.com)

N. Achour is Université des sciences et de la technologie Houari Boumediene Algiers, Algeria. (e-mail : noura.achour@gmail.com).

[12] and FAB-MAP2 [13] demonstrated successful scene recognition in large scale environments. The principal disadvantages of this method lies in the need of an offline process for dictionary building and the computation cost for both features extraction and matching search. A good dictionary for one environment can be useless in other environments as it has been demonstrated by different papers [18][38][39][28]. The core idea of FAB-MAP has been used by Glover et al. [14] to perform the visual data association for RatSLAM approach. But, no significant advantage of this combination over the RatSLAM-only system has been observed. The obtained results showed some promises of FAB-MAP's ability for scene recognition for different lighting conditions.

Kawewong et al. [18] proposed the Position-Invariant Robust Features (PIRF) for scene recognition in dynamic outdoor environments. Scene recognition proceeds by matching individual feature to a set of features of testing images and identifying the place with the highest matched features. Scene recognition using this method achieves a high recall rate at 100% precision. However, it is computationally expensive and uses a large amount of memory in the redundant process of keeping signatures of places.

Bacca et al. [2] build a feature management system known by Feature Stability Histogram (FSH) for appearance-based mapping and localization. The proposed feature management system is able to cope with changing environments and long-term operation. The appearance of the indoor environments is described using SURF descriptors and it is stored using an FSH at each node in the robot topological map.

The global image descriptors aim to represent an image using a global information such as intensity profile [27][39], color histograms[38], Gist [42] and others.

An interesting place recognition system described by Siagian et al. [42] uses models of human vision to extract both salient features (such as color, intensity, etc.) and the Gist of the scene to form a scene descriptor. In Sünderhauf and Protzel [44], a Binary Robust Independent Elementary Features (BRIEF)-Gist has been proposed.

All the cited methods share the same main idea that consists in the extraction of image features and their saving into a database or a dictionary building. The extracted features of the current image are used to decide whether it has already perceived at some time in the past or whether this scene represents a new and unknown area in the environment. However, a drawback of these approaches is that they search matching between current views and the stored data, which involves expensive computation and storage costs.

Today, Researches in the field of place and scene recognition for mobile robots have focused their investigations on biologicaly inspired mechanisms [26][3][17][2][30]. The proposed solutions use the results of human and animal space cognition systems study. This problem is generally posed as two questions: first, how do the human and animals brain build a cognitive representation of their environments and how to recognize places? and, second, how to develop the best emulated models of the brain to solve robotics problems?

The rest of this section discusses a number of leading bio-inspired architectures for place recognition. Milford and Wyeth [27], Milford [26], and Sünderhauf and Protzel [43] investigate the usefulness of bio-inspired computational model oriented for robotic applications and especially localization and mapping in real, large and complex outdoor environments. These strategies are based on the Place Cells (PCs) properties discovered in the rat's brain. In these works, the place field size is considered fix, which is not a real characteristic of PCs [35], and the number of the used neurons for space cognition is considered known. The local View Cells (VCs) save all the extracted visual information during robot motion. Like other appearance-based place recognition approaches, all the saved information are used for previously visited place recognition.

Various mobile robot architectures with spatial cognition have been proposed in [3][17]. Despite the fact that these approaches capture some properties of the rat's brain such as the PCs features, they consider only the visual information derived from artificial landmarks. These approaches have been tested only in very small environments (maze environments). Therefore, their usefulness in outdoor environment needs more investigations.

Motivated by the neurobiological features of the human memory, Nguyen et al. [33] have proposed recently a new spatio-temporel memory architecture to solve the problem of semantic visual place recognition for mobile robot localization in indoor environments. The KFLANN (an ART based neural network) has been used as the core unit in the visual scene quantization. While this approach contains many valuable ideas, it has not yet been evaluated for large-scale and real outdoor application.

The approach proposed by Becca et al. [2] is based on a model of the human memory. The extracted visual descriptors are classified as either stable or non-stable using Long-Term (LTM) and Short-Term Memory (STM) mechanisms.

The review of related works to the appearance-based place recognition approaches shows the use of simple or omnidirectional cameras as the main sensor. However, actually many researches use sensor fusion in order to complement the appearance-based model with odometric [22], range [32] and thermal [25] measures. These approaches take advantage of other metrics more robust to a variety of illumination and environmental changes such as infrared thermal image [25]. Other approaches incorporate the sequential nature [28][33][36] of the observed scenes in place recognition to overcome the perceptual aliasing problem.

In our previous papers [38][39], a computational model of the visual memory using Fuzzy ART network has been proposed. This approach has been inspired from view cells properties. The visual memory building is done with incremental, competitive and unsupervised learning. The scene recognition is carried out without saving the extracted

visual information and without searching explicitly similarity between current and previous scenes. According to how the input of the Fuzzy ART is employed for searching the winning view cell, the image descriptor should be of a fixed size and composed of ordered components. The purpose of the current paper is to propose and test the use of other image descriptor to improve the recognition quality.

III. Visual Memory Model Using The ART Theory

This section describes the pipeline process of the proposed approach. For that purpose, we begin with reviewing the core ideas of Adaptive Resonance Theory. Then, our previous model of the visual memory based on Fuzzy ART [9] is summarized.

A. *ART theory*

The Adaptive Resonance Theory (ART) has been proposed by Carpenter et al. [6][7] to model how the brain autonomously learns to categorize, recognize, and predict objects and events in a changing world. This theory has been introduced to overcome the so-called stability-plasticity dilemma [7]. This dilemma describes the brains capacity of learning quickly new information without causing catastrophic forgetting. A large variety of the ART architectures has been developed: ART1, ART2, ART3[6], Fuzzy ART[9] and Fuzzy ARTMAP[8] bART[10] and ARTSCENE[15]. The training algorithms of the ART networks is incremental and include both unsupervised and supervised learning conditions. To date, there are many studies and successful applications of ART networks in different areas.The recently published paper of Grossberg [16] presents a review of classical and recent development of ART.

The Fuzzy ART network [9] is a kind of ART neural network for competitive learning with winner-take all coding. This network has unique advantages and characteristics that lead to be a very good candidate for visual memory modeling. The attractive features of this neural network like the possibility of fast, online, incremental, and unsupervised learning can be exploited to develop the visual memory model and solve the scene recognition problem.

B. *The visual memory model*

The schematic diagram of the proposed model of the visual memory is shown in Fig. 1. It is composed of two modules: the vision module and the visual memory module. The first module extracts the image descriptor and transmits it to the visual memory as it will be detailed in section 4. The goal of the last module is to build, incrementally, a cognitive representation of the environment and to recognize if the current scene has been previously observed or it is a new one.

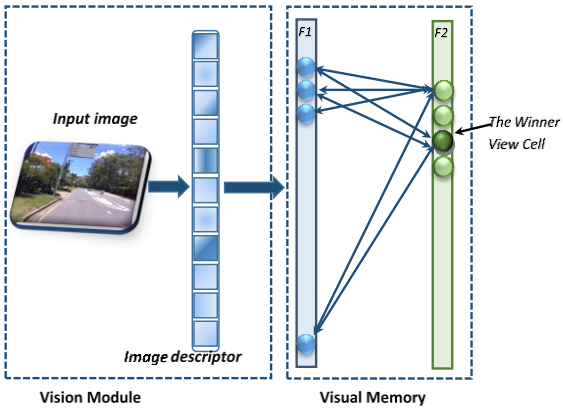

Figure 1. The schematic diagram of the visual memory

C. *The principle of the visual memory based on Fuzzy ART*

The experimental study of space cognition in rodent's brain reported the presence of Place Cells (PCs) in the hippocampus [35]. The firing of these cells is maximal when the rodent is located at a specific location in the environment. Many neurobiologists concentrate their research on the study of PCs firing properties and the development of different computational models [35][26][17]. Solving robotics problems by inspiration from the neurobiology has been the subject of a large body of research over the last years. RatSLAM [26][27] is one of the most successful bio-inspired SLAM approach.

Place cells have also been described in primate's hippocampus including humans, as have a variety of other spatial cells: Head Direction cells (HD) and spatial View Cells (VC) that respond to the location at which the animal is looking. Rolls and colleagues [40] have discovered these cells in primate's hippocampus. The experiments carried out by these researches show the increased firing rate of these cells when the animal looked at a particular part of the environment. The existence of spatial-view cells might be an indication that primates have the ability to identify and recognize places and their contents without physically visiting those places. The properties of these cells displayed "visual memory" features such as object and scene recognition. This memory has an important role in the spatial mapping system.

Besides the attractive features of the view cells, the Fuzzy ART, as previously mentioned, is a good candidate for the design of a visual memory in order to solve the problems of scene recognition and loop closure detection by mobile robot.

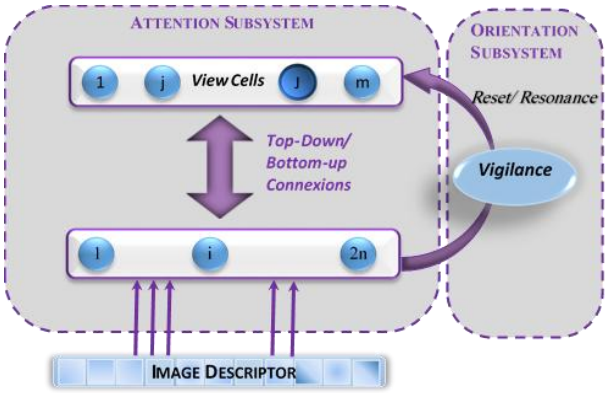

Figure 2. Fuzzy ART based visual memory

The proposed visual memory [39] has the architecture of the Fuzzy ART neural network (Fig. 2). For each perceived scene, an input vector is generated by the vision module (Fig. 1) to describe its appearance into a specific intentional region. The output layer of the Fuzzy ART network consists of view cells. To start the incremental building of the visual memory, one view cell is created. Then, in presence of another input vector, the orientation subsystem decides if it is a new place with a new appearance and creates a new VC representing this view or it is a previously observed scene. In this case, an old VC fires maximally using the winner takes all rule. The single active neuron (VC) corresponds to the closest previously learned appearance. The re-activated view cell indicates the recognition of previously visited place and the detection of a closed loop. The detailed algorithm can be found in the previous papers [38][39].

IV. LOCAL HS COMBINED HISTOGRAM

The study of the primate's and human's vision system has shown its sensitivity to the global appearance of a scene without the need of object centered information for rapid scene recognition [15][33][40]. In the scene-centered approaches, the high-level scene recognition uses a description of the entire scene. Moreover, other neurobiological studies reveal that the primates use a visual attention process to fix their attention in the perceived scene on particular regions [15]. The visual attention mechanism consists on processing only the interesting part of the visual information and ignoring what does not; this is the main goal of visual attention.

Because of the cited observations about human's vision system, the global image descriptor proposed in this paper consists on computing a global descriptor within a visual attention region. For outdoor urban application, the visual attention (VATT) window is located up of the horizon line in order to eliminate the road pixels and reduce the effect of the presence of moving objects such as cars and pedestrians.

A. Color space

The HSV color space reflect human perception and identification of color images. It is more intuitive than the RGB color model. In this space, the hue component (H) represents the color tone (for example, red or blue), the saturation (S) is the amount of color (for example, bright red or pale red). The third component value (V) is the amount of light. Because this last component is very sensitive to lighting condition, the proposed descriptor for scene recognition will not use this component. In our previous work [38], the histogram representing the HS component is used where the two components histograms are computed independently. The obtained results showed the need of improving this descriptor. The aim of this paper is to build a local HS combined histogram as a global image descriptor. The main disadvantage of the color histogram for scene recognition and similarity measure is the lack information about the spatial distribution of colors within the image. To overcome this problem, the VATT region is downsampled to equal patches like the procedure used for Gist computation [44]. The determination of color histograms within local regions leads to descriptor robust to small variation in the camera pose and orientation. The proposed HS combined histogram describes the visual appearance of places in a compact manner and its computation is both easy and fast.

B. Building the HS combined histogram

Usually, a color histogram in RGB color space is computed by calculating an N-bin histogram for each of the R, G and B color bands. However, considering the HSV color space, the human visual system is more sensitive to the hue component (H) than to the saturation one. The H component is quantized finer than the S component. We can said that we can distinguish for example eight (NH=8) tones with four (NS=4) levels of saturation. It is worth noting here that the color of one pixel is a combination of the two information (H and S). In our previous work [38], the HS histograms are computed independently for the two components. Unfortunately, this hypothesis (used also for RGB histogram computation) loses the 2D spatial information of the HS pairs in this color space. In order to retain the 2D spatial information, the combined HS histogram consists of NH*NS bins. Using NH=8 and NS=4 keeps the descriptor computation fast. Therefore, 32 colors are considered in HS combined histogram. Downsampling the VATT window into 4 patches leads to describe an image by a vector of 128 elements. In comparison, a single standard SIFT feature is represented through a 128 dimensional vector and a single image can be described by thousands of SIFT features.

V. RESULTS

The performance of the proposed algorithms was evaluated in urban environments. The dataset (st_lucia), provided by Glover et al. [14] was chosen to test the proposed fuzzy ART-based visual memory and the local HS combined histogram in large scale outdoor environment. This dataset is composed of both visual and GPS data obtained from a car driven around suburban streets in St-Lucia, Queensland (Fig 3).

This paper presents our primary results. Future studies will be focused on testing the scene recognition performance faced with severe environmental changes.

The purpose of this experiment is to evaluate how the proposed visual memory is able to recognize previously observed scenes during the vehicle motion. Fig. 4 illustrates the correctly detected closed loops at 100% precision while Fig. 5 shows the whole returned false positive at the experiment end. The precision-recall plot (Fig. 6) shows an acceptable recall ratio at 100% precision (78%). Note that the proposed approach gives a binary decision of scene recognition; the precision-recall curve is obtained by interpolation of the recall and precision values at different points. These points are chosen by the evaluation of the distance between locations (using GPS measurements) where the same view cells are activated according to a fixed threshold.

However, the returned false positive imposes future investigation by the introduction of visual memory correction procedure that considers the sequential nature of the observed scenes during the robot motion.

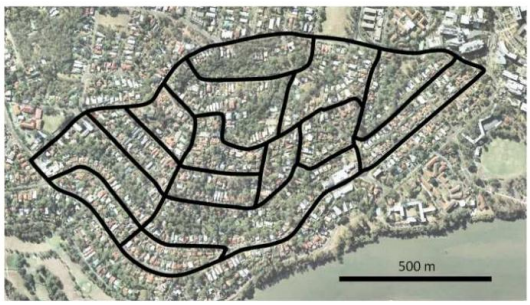

Figure 3. The aerial photograph and the ground truth data obtained from [14]

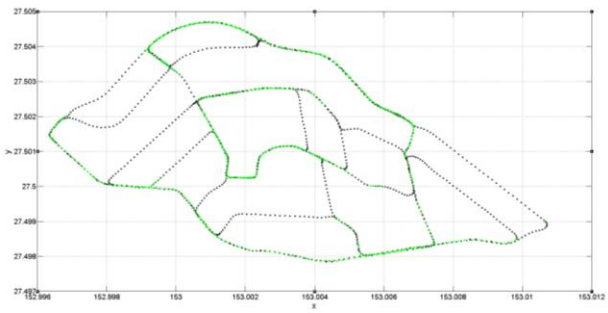

Figure 4. Detected loop closures (green) at 100% precision imposed on the ground truth

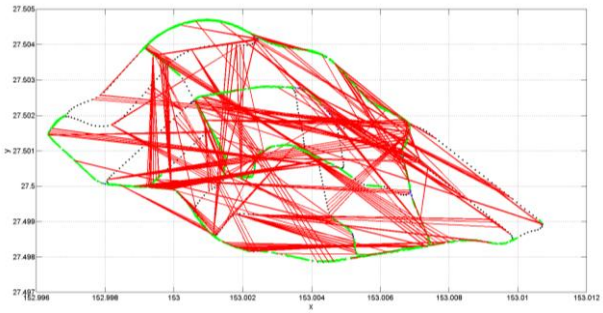

Figure 5. All the returned false positives at the experiment end

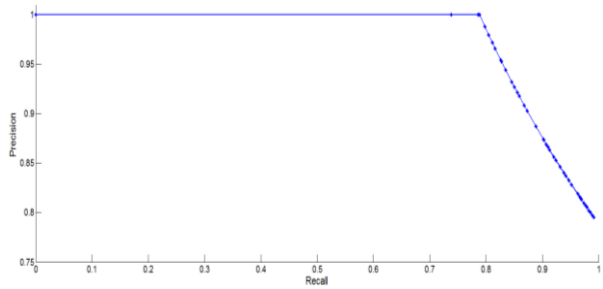

Figure 6. The Precision-recall curve

VI. Conclusion and Future Works

This paper concentrates on scene recognition problem for mobile robot place recognition and loop closure detection. The proposed solution aims to build a neurobiological inspired computational model of the visual memory. The proposed system is composed of two subsystems including vision module and visual memory. The first extracts a simple and efficient global image descriptor while the second utilizes a Fuzzy ART neural network composed of view cells for scene recognition. The global image descriptor extracted from the HS enhanced image presents a very interesting features such as robustness face to illumination change, it is compact and easy to compute.

The proposed approach has been tested using a large-scale outdoor datasets in changing environment. The results showed that the proposed HS combined histogram can describe efficiently the appearance of an outdoor scene. The fuzzy ART-based visual memory can recognize quickly a previously observed scene without the need of saving the extracted visual information. The Fuzzy ART connections weights save the signature of each scene. The scene recognition is carried out via the pattern resonance within the visual memory without requiring an explicit step for the search of similarity between frames.

This work serves as a one-step towards the development of a bio-inspired model of the visual memory. The exploitation of this type of memory allows mobile robot to build a cognitive representation of its environment; it can recognize and remember previously acquired information.

Experiments revealed a weakness of the proposed approach based only on visual information. Future work will

focus on how considering the sequential aspect of the images captured incrementally during the robot motion in the development of the visual memory. The exploitation of the temporal information seems to be a very interesting solution to overcome the problem of perceptual aliasing. Therefore, the correction of the visual memory and of its outputs (new or recognized scenes) over time need to be considered especially for lifelong operation in changing environments.

REFERENCES

[1] Angeli A, Filliat D, Doncieux S, and Meyer J.-A (2008), "A Fast and incremental method for loop-closure detection using bags of visual words," IEEE Tran. on Robotics 24(5), pp. 1027-1037.

[2] Bacca B, Salvi J and Cufí X (2011), "Appearance-based mapping and localization for mobile robots using a feature stability histogram," Robotics and Autonomous Systems 59 (2011) 840–857

[3] Barrera A and Weitzenfeld A (2008), "Biologically-inspired Robot Spatial Cognition based on Rat Neurophysiological Studies," Journal of Autonomous Robots, Vol. 25 Issue 1-2, pp. 147-169

[4] Callmer J, Granström K, Nieto J and Ramos F (2008), "Tree of Words for Visual Loop Closure Detection in Urban SLAM," 2008 Australasian Conf. on Robotics & Automation.

[5] Calonder M, Lepetit V, Ozuysal M, Trzcinski T, Strecha C, and Fua P (2012), "Brief: Computing a local binary descriptor very fast," IEEE Trans. Pattern Anal. Mach. Intell., 34(7):1281–1298, July 2012.

[6] Carpenter G. A and Grossberg S (1990), "ART 3: hierarchical search using chemical transmitters in self- organizing pattern recognition architectures. Neural," Networks, 3, 129–152

[7] Carpenter G, Grossberg S (1987) A massively parallel architecture for a self-organizing neural pattern recognition machine. Comput Vis Graph Image Process 37:54–115

[8] Carpenter G. A, Grossberg S, Markuzon N, Reynolds J. H and Rosen D. B. (1992), "Fuzzy ARTMAP: a neural network architecture for incremental supervised learning of analog multidimensional maps," IEEE Transactions on Neural Networks, 3, 698–713.

[9] Carpenter G. A, Grossberg S and Rosen D. B (1991), "Fuzzy ART: fast stable learning and categorization of analog patterns by an adaptive resonance system," Neural Networks, 4, 759–771

[10] Carpenter G. A and S. C Gaddam, S. C. (2010), "Biased ART: a neural architecture that shifts attention toward previously disregarded features following an incorrect prediction," Neural Networks, 23, 435–451.

[11] Cummins M and Newman P (2008), "FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance," Int. Journal of Robotics Research, vol.27, pp. 647-665.

[12] Cummins M and Newman P, (2010), "Accelerating FAB-MAP With Concentration Inequalities," IEEE Transactions on Robotics, Vol. 26, No. 6, pp. 1042-1050.

[13] Cummins M and Newman P (2011), "Appearance-only SLAM at large scale with FAB-MAP 2.0," Int. Journal of Robotics Research, Vol. 30 Issue 9, pp. 1100-1123.

[14] Glover A, Maddern W, Milford M, Wyeth G (2010), "FAB-MAP + RatSLAM: Appearance-based SLAM for multiple times of day," IEEE Int. Conf. on Robotics and Automation (ICRA), pp. 3507-3512.

[15] Grossberg S and Huang T. R (2009), "ARTSCENE: a neural system for natural scene classification," Journal of Vision, 9(6), 1–19

[16] Grossberg S (2013), "Adaptive Resonance Theory: How a brain learns to consciously attend, learn, and recognize a changing world", Neural Networks 37 (2013) 1–47

[17] Hasson C, Gaussier P (2010), "Path Integration Working Memory for multitasks dead reckoning and Visual Navigation," Pro, of the 11th Int. Conf. on Simulation of Adaptive Behavior, SAB 2010, pp. 380-389, Paris, France

[18] Kawewong A, Tongprasit N, Tangruamsub S and Hasegawa O (2011), "On line and Incremental Appearance-based SLAM in Highly Dynamic Environments," Int. Journal of Robotics Research archive Vol. 30 Issue 1, pp 33-55.

[19] Konolige K, Bowman J, Chen J, Mihelich P and Garage W (2010), "View-Based Maps," Int. Journal of Robotic Research, pp. 941-957.

[20] Kumar A, Tardif J, Anati R and Daniilidis K, (2008), "Experiments On Visual Loop Closing Using Vocabulary Trees," In Computer Vision and Pattern Recognition Workshops, (CVPRW), Anchorage, AK.

[21] Leutenegger S, Chli M, and Siegwart R (2011), "BRISK: Binary robust invariant scalable keypoints," In Proceedings of the IEEE International Conference on Computer Vision, 2011.

[22] Lowry S. M, Wyeth G. F, Milford M (2013), "Odometry-driven Inference to Link Multiple Exemplars of a Location," IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 534 – 539, Tokyo

[23] Lui W. D and Jarvis R (2010), "A Pure Vision-based Approach to Topological SLAM," IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS), pp. 3784-3791, Taipei, Taiwan.

[24] Lowe D. G (2004), "Distinctive image features from scale-invariant keypoints," Int. J. Comput. Vision, 60(2):91–110, November 2004.

[25] Maddern W and Vidas S (2012), "Towards Robust Night and Day Place Recognition using Visible and Thermal Imaging," Proc. of Robotics Science and Systems (RSS): Workshop on Alternative Sensing Modalities

[26] Milford M (2008), "Robot Navigation from Nature," vol. 41, Berlin-Heidelberg: Springer-Verlag.

[27] Milford M and Wyeth G (2008), "Mapping a suburb with a single camera using a biologically inspired SLAM system," IEEE Transaction on Robotics, Vol. 24, pp: 1038-1053.

[28] Milford M and Wyeth G (2012), "SeqSLAM: Visual Route-Based Navigation for Sunny Summer Days and Stormy Winter Nights," Proc, of the IEEE Int. Conf. on Robotics and Automation (ICRA), pp. 1643-1649, Saint Paul, MN

[29] Milford M and Wyeth G (2011), "Improving recall in appearance-based visual SLAM using visual expectation," Proc. Australian Robotics & Automation Association (ACRA), Brisbane, QLD.

[30] Milford M and Jacobson A (2013), "Brain-inspired Sensor Fusion for Navigating Robots," Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA), pp. 2906-2913

[31] Newman P and Ho K (2005), "SLAM- Loop Closing with Visually Salient Features," Int. Conf. on Robotics and Automation, Barcelona, Spain.

[32] Newman P, Cole D and Ho K (2006), "Outdoor slam using visual appearance and laser ranging," Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA), Orlando, Florida, USA.

[33] Nguyen V. A, Janusz A. Starzyk J. A and Goh W (2013), "A spatio-temporal Long-term Memory approach for visual place recognition in mobile robotic navigation", Robotics and Autonomous Systems 61 (2013) 1744–1758

[34] Nister D and Stewenius H (2006), "Scalable recognition with a vocabulary tree," Proc. of the Conf. on Computer Vision and Pattern Recognition, Vol. 2, pp. 2161–2168.

[35] O'Keefe J and Nadel L (1978). "The hippocampus as a cognitive map," Oxford University Press.

[36] Pepperell E, Corke P and Milford M (2013), "Towards Persistent Visual Navigation using SMART," Proc. of Australasian Conference on Robotics and Automation, Sydney Australia

[37] Redish A (1997), "Beyond the cognitive map," Ph.D. thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

[38] Rebai k, Azouaoui O and Achour N (2012), "Bio-Inspired Visual Memory for Robot Cognitive Map Building and Scene Recognition," IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS 2012), Vilamoura, Portugal.

[39] Rebai K, Azouaoui O and Achour N (2013), “Fuzzy ART-based place recognition for visual loop closure detection,” Biological Cybernetics (2013) 107:247–259

[40] Rolls E (1999), “Spatial View Cells and the Representation of Place in the Primate Hippocampus,” Journal of Hippocampus Vol. 9, pp. 467-480.

[41] Rublee E, Rabaud V, Konolige K, and Bradski G (2011), “Orb: An efficient alternative to sift or surf,” In International Conference on Computer Vision, Barcelona, 11/2011

[42] Siagian C and Itti L (2009), “Biologically Inspired Mobile Robot Vision Localization,” IEEE Tran. on Robotics, Vol. 25, pp. 861 – 873.

[43] Sünderhauf N and Protzel P (2010), “From Neurons to Robots: Towards Efficient Biologically Inspired Filtering and SLAM,” *Advances in Artificial Intelligence*, Lecture Notes in Computer Science, 6359, pp. 341-348, Springer Verlag.

[44] Sünderhauf and Protzel (2011), “BRIEF-Gist - Closing the loop by simple means,” IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS), pp. 1234-1241, San Francisco, California, USA.

[45] Ulrich I and Nourbakhsh I (2000), “Appearance-Based Place Recognition for Topological Localization,” IEEE Int. Conf. on Robotics and Automation (ICRA), pp. 1023-1029, San Francisco, California, USA.