A Biological Inspired Visual Landmark Recognition Architecture

Quoc Do and Lakhmi Jain

University of South Australia,
Mawson Lakes Campus, South Australia, Australia
Email: {Quoc.Do, Lakmi.Jain}@unisa.edu.au

Abstract - **An architecture that is inspired by a human's capability to autonomously navigate an environment based on visual landmark recognition is presented. It consists of pre-attentive and attentive stages that allow visual landmarks to be recognized reliably under both clean and cluttered backgrounds. The pre-attentive stage provides an efficient means for real-time image processing by selectively focusing on regions of interest within input images. The attentive stage has a memory feedback modulation mechanism that allows visual knowledge of landmarks in the memory to interact and guide different stages in the architecture for efficient feature extraction and landmark recognition. The results show that the architecture is able to reliably recognise both occluded and non-occluded visual landmarks in complex backgrounds.**

Keywords: visual landmark recognition; neural network; image processing

I. INTRODUCTION

Human vision has an exceptional ability to perceive the external environment. Therefore, it is highly desirable to mimic this ability in an autonomous robot. However, this requires an in-depth knowledge of the human vision system in terms of behavioural, structural and computational mechanisms. Fortunately, advances in the fields of biology, neurophysiology and cognitive neuroscience has enhanced the understanding of various aspects of visual processing [1-4], leading to the development of artificial neural networks that model biological neural processing.

Adaptive Resonance Theory (ART) was first introduced as a physical theory of cognitive information processing in the brain [5, 6], and it is derived from a simple feedforward real-time competitive learning system called Instar [7], addressing the *plasticity-stability dilemma* wherein a real-time competitive learning system must be *plastic* to incorporate new and significantly novel events, while being *stable* to avoid the corruption of previously learned information in the memories by erroneous observations. Since the introduction of ART in the late 1970's and early 1980's, a large family of ART-based artificial neural network architectures had been developed by Grossberg and Carpenter, which include: ART-1 for binary inputs [8], ART-2 for binary and analog inputs [9], ART-3 for hierarchical neural architectures [10].

The ART neural networks have gained popularity in various engineering applications, solving many non-linear problems, and text and image classification problems. However, they have a key deficiency which prevents them from recognising a familiar pattern embedded in a cluttered background [11].

The recognition of objects in complex backgrounds is a difficult task, primarily due to parts of the object being merged or occluded by other background features. In contrast to the ART networks, it has been reported in [11, 12], that for the networks to recognise a familiar object in a cluttered environment requires prior image segmentation. However, such preprocessing diminishes the role of neural architectures in object recognition. Fortunately, studies in neurophysiology have suggested that visual attention has modulatory effects on neuronal signals [13, 14] and top-down mechanisms from memory may influence the activations of the desired bottom-up stimuli [15-17] in contrast with Ulman's proposal [18], that the top-down feedback connection is directly involved in the activation of the lower neural layer. This leads to an interesting suggestion that the feedback connection from the higher cortex modulates the input features to be attended to or ignored and this has been supported by evidence produced from many studies [19, 20]. In 1997, Lozo was inspired by these findings and developed the SAART network [11]. It is an extension to the ART-3 network by incorporating top-down feedback pathways to modulate the bottom-up input pattern, and enables the network to selectively remove background features so that it can recognise an object embedded in a complex scene.

The SAART neural network is well known for its ability to achieve memory visual resonance, which removes relevant and noisy data from the input image, and enables landmark recognition in cluttered backgrounds. However, it is subjected to a major drawback. As reported in [7], it is a computationally intensive dynamic network, thus not suitable for real-time landmark or object recognition applications. Furthermore, the target of interest is assumed to be completely visible. In contrast, objects or landmarks in the natural environment are subjected to varying levels of occlusion by adjacent features. This paper describes an extension to the SAART neural network that enables the

978-0-7695-3866-2/09 $26.00 © 2009 IEEE
DOI 10.1109/DICTA.2009.61

architecture to cope with the recognition of both partially occluded and non-occluded landmarks in real-time.

II. Image Processing Architecture

The proposed architecture provides a means for both visual knowledge (stores as memory templates) to facilitate feature extraction – *top-down modulation*, and simultaneously allows selected features to facilitate the active memory template – *bottom-up modulation*. The convergence of these bottom-up and top-down memory interactions enables recognition of partially occluded and non occluded landmarks in cluttered backgrounds. The architecture consists of two stages: pre-attentive and attentive that are inspired by the human vision system and is illustrated in Figure 1. The former focuses on rapid identification of regions of interest (ROIs) within input images prior to any intensive processing, which significantly reduces the image processing time while the latter concentrates on feature extraction and landmark recognition within the ROIs. Both of these stages use the Memory Feedback Modulation (MFM) mechanism for facilitation and inhibition of relevant and irrelevant visual information respectively.

A. Attentive Stage

1) Memory Database

The architecture recognises visual landmarks using template matching of an input extracted pattern with an active memory template. There are two types of memory templates: memory image and binary memory filters. The former is used in the landmark recognition stage for object classification, while the latter is used to create memory feedback modulation pathways in the MFM mechanism, which provides memory guidance for the feature extraction, searching and matching stages.

Each memory template is used to create three binary memory filters. Firstly, the template is further divided into twenty five smaller sub-memory templates (SMTs) as illustrated in Figure 2(b). The first filter is created using a 5x5 array. Each cell in the array indicates the status of a STM. This filter is called the memory active region (MAR) filter, as shown in Figure 2 (d). The second filter is named the memory active edge (MAE) filter. It is created by analysing active edge distribution within each SMT using *eq.1*. The MAE filter has the same size as the memory template, where each high pixel denotes a corresponding active pixel in the memory template, as illustrated in Figure 2 (c). Both MAR and the MAE filters are used to provide memory guidance for the feature extraction, and searching stages. The third filter is named the landmark enclosed area (LEA), with the same size as the memory template, and denotes the entire region that encloses the landmark as shown in Figure 2 (e). This filter is used in the template matching stage.

$$F(i,j) = \begin{cases} E(i,j) > \tau & F(i,j) = 1 \\ E(i,j) < \tau & F(i,j) = 0 \end{cases} \quad \ldots (1)$$

Where E(i,j) is the edge processed image, τ is a small threshold and F(i,j) is the memory active edge (MAE) filter.

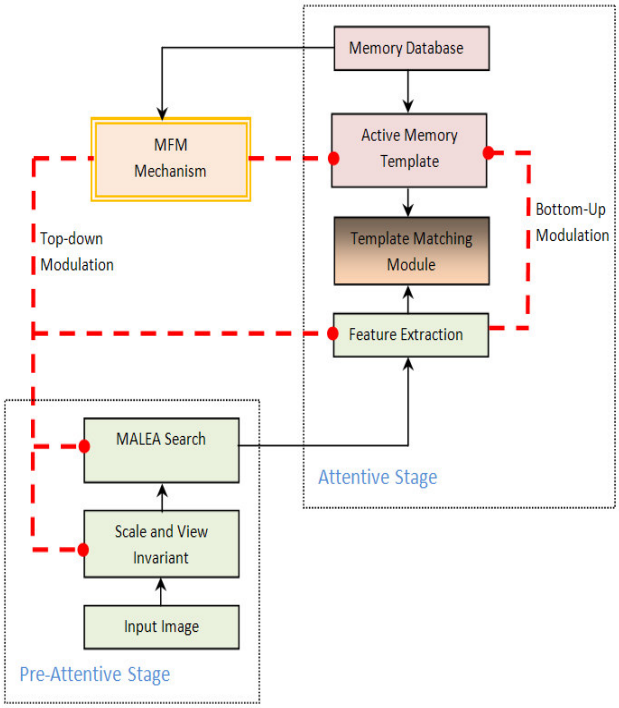

Figure 1. The biological visual landmark recognition architecture that combines pre-attentive and attentive stages from the human visual system, and memory feedback modulation from the SAART neural network.

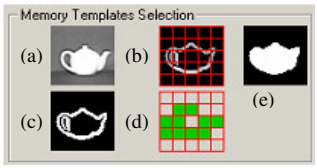

Figure 2. A memory template and the corresponding binary memory filters. (a) The gray-level image, (b) the memory template, (c) the memory active edge (MAE) filter, (d) the memory active region (MAR) filter and (e) landmark enclosed area (LEA) filter.

2) Memory Feedback Modulation Mechanism

The prominent characteristic of the architecture lies within the incorporation of the memory feedback modulation (MFM) mechanism, which enables the architecture to selectively attend to relevant input data, while ignoring irrelevant visual information. As a result, this enables the architecture to reliably achieve object-background separation, which leads to the ability to recognise visual

landmarks in cluttered backgrounds. For instance, consider an input image entering the feature extraction stage, where it is modulated by the MFM mechanism as illustrated in Figure 3.

The MFM mechanism allows prior visual knowledge from the memory template to selectively guide the feature extraction process. This is achieved by applying *eq.2* to the input. Notice that the memory template has approximately 20% of high pixel values that define the shape of the landmark. These pixels form memory feedback pathways governed by the MAE filter, creating amplification channels. Pixels that correspond to these channels are amplified or otherwise passed through unaffected. The modulated region, P(i,j) is subjected to lateral completion by L2 normalisation, resulting in lateral suppression that removes irrelevant visual information within the input region via a threshold. Thus it achieves landmark-background separation.

$$P(i,j) = ROI(i,j)\left[1 + G * BMF(i,j)\right] \quad \ldots (2)$$

Where P(i,j) is the result of the memory modulation, ROI(i,j) is the region of interest in the input image, BMF(i,j) is the binary memory filter and G is a gain control.

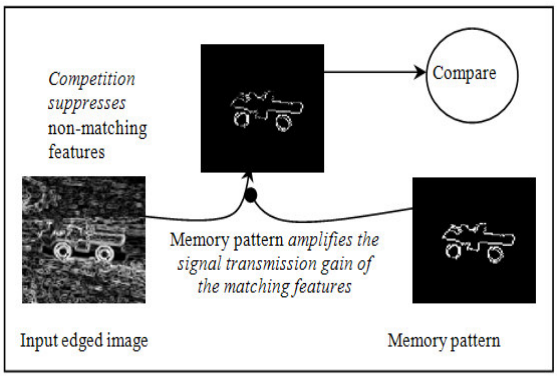

Figure 3: The memory feedback pre-synaptic facilitation and inhibition for selectively enhancing desired features while suppressing background clutters.

3) *Landmark Recognition*

The landmark recognition has two modules: matching and evaluation. The former uses the MFM mechanism to generate governing matching channels using the MAR filter, the MAE filter, and an additional dynamic filter called the landmark occluded area (LOA) filter. The LOA filter is dynamically determined at the feature extraction stage. It is a 5x5 filter with each cell indicating an area (10x10 pixels) as being a potential landmark occlusion. The LOA filter is determined in a similar way to the MAR filter as described in section 2.1.1. Then, the matching channels, $C_k(i,j)$ are created using *eq.3*. Only pixels that lie within the channels are considered in the matching process.

$$C_k(i,j) = MAE_k(i,j)\left[MAR_k(i,j) - LOA_k(i,j)\right] \quad \ldots (3)$$

Where $C_k(i,j)$ is the matching channel for each active ROI patch, MAR(i,j) is the memory active region filter, LOA(i,j) is the landmark occluded area filter and $MAE_k(i,j)$ is the corresponding region in the memory active edge filter.

The template matching process performs a similarity measure between each active STM and the corresponding patch in the input ROI region – denoted as a degree of match (DoM). The matching process is expressed mathematically in *eq.4*. Each DoM has a value range from 0 to 1, where 1 represent 100% match. This is further evaluated against a matching threshold, where the input patch with a DoM value greater than the threshold is regarded as a STM match. If the summation of all the number of STM matches within the ROI region is greater an occlusion threshold, then this region is passed into the evaluation module for further validation to ensure robust landmark recognition.

$$DoM = \frac{\sum ROI(i,j)M(i,j)C_k(i,j))}{\varepsilon + \sqrt{\sum (C_k(i,j)ROI(i,j))^2}\ \sqrt{\sum (C_k(i,j)M(i,j))^2}} \quad .. (4)$$

Where ROI(i,j) is the patch that corresponds to the active STM and M(i,j) is the STM, ε is a small constant to prevent the equation from being divided by zero.

The evaluation module validates each positive match by assessing the similarity between the regions that enclose the landmark in the memory template - denoted by the LEA filter, and the corresponding area in ROI region, but excluding the detected occluded areas - denoted by the LOA filter. This is achieved by creating evaluation channels using *eq.5*. Similarly, the evaluation process focuses on pixels in the input ROI and the memory template that lie within the evaluating channel, while ignoring all others. The matching validation is achieved using *eq.4*, with $C_k(i,j)$ replaced by $E_k(i,j)$. The overall DoM is calculated by averaging all of the DoM values. This is measured against an evaluation threshold, and a match is declared for the DoM value greater than the threshold.

$$E_k(i,j) = LEA_k(i,j)\left[MAR_k(i,j) - LOA_k(i,j)\right] \quad \ldots (5)$$

Where $E_k(i,j)$ is the evaluating channel for each active ROI patch, MAR(i,j) is the memory active region filter, LOA(i,j) is the landmark occluded area filter and $LEA_k(i,j)$ is the landmark enclosed area filter.

B. Pre-Attentive Stage

1) Memory Assisted Local Edge Analysis

The central idea in the memory assisted local edge analysis (MALEA) approach method is the incorporation of the

MFM mechanism, which involves the use of the memory active edge (MAE) and the landmark enclosed area (LEA) filters to provide memory guidance for determining the ROIs within the input image. The MAE and LEA filters are described in section 2.1.1. The MALEA search considers pixels that correspond to the MAE and LEA memory filters, as only these edges are relevant in describing the shape of the landmark. All others can be safely discarded, which significantly reduces the amount of computation required.

The operation of the MALEA search is illustrated in Figure 4**.** Initially, input patches are extracted as the search window scans across the image. The regions that satisfy the ROI-threshold are passed through and further evaluated with the signature threshold to be confirmed and classified as ROIs, otherwise these regions are discarded. This process is guided by the visual knowledge feedback from memory using the MAE and LEA filters. Both the ROI and the signature thresholds are set dynamically for every active memory template.

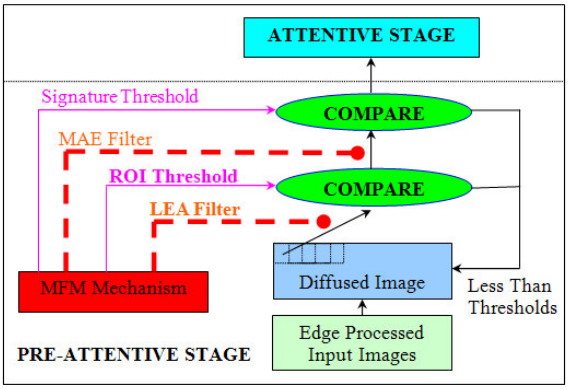

Figure 4. The process of the MALEA method, which employed the MFM mechanism (using the MAE and the LEA filters) to provide memory-guided selective processing for rapid determination of ROIs within an input image.

2) Scale View Invariant Landmark Recognition

The shape of a landmark in an input image is affected by noises and 3D to 2D projection, image distortion, and small size and shape changes. These result in faulty landmark recognition due to template mismatch as the MFM mechanism requires elementary alignment of input data with the memory feedback pathways.

In order to overcome image distortions, and small changes in size and shape, the developed architecture employs two concepts named band transformation and shape attraction [11]. The central idea is that, if an edge is missing due to distortion or small size and shape changes, it can be compensated for or reconfirmed by considering its neighborhood. This method has two parts. Firstly, band transformations is applied to diffuse the shape of the input pattern by mean of a Gaussian filter or an averaging mask [21] to produce a diffused-edge image. Secondly, the distorted information is recovered by applying a memory guided shape attraction process to the diffused image. The shape attraction process uses the MFM mechanism to selectively attract and recover the missing edges. The concept of image diffusion and shape attraction processes to achieve distortion and small size and shape invariant recognition is further illustrated in Figure 5.

The concepts of band transformation and memory guided shape attraction affectively achieve memory driven feature extraction. These concepts have been used to develop a method named simultaneous multiple-memory image search (SMIS) for achieving size and view invariant landmark recognition [22].

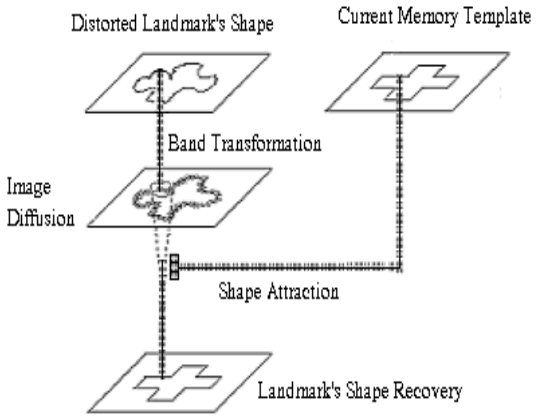

Figure 5. The memory-guided shape attraction process using the MFM mechanism to selectively project and recover missing or distorted input edge information.

III. RESULTS

The biologically inspired visual landmark recognition architecture described has been successfully evaluated using images from both clean and complex backgrounds, consisting of occluded and non-occluded situations.

A. Non-Occluded Landmark Recognition

In order to evaluate the architecture's capability in recognizing visual landmarks in cluttered backgrounds, a comparison study with a traditional template matching approach was conducted. A total of forty different input images were collected, composed of twenty in clean-backgrounds and twenty in cluttered backgrounds. These images were fed both into the proposed architecture, and the traditional template matching method to evaluate and compare their performances. All image processing stages were kept constant, the only difference is the additional MFM mechanism incorporated in the proposed architecture.

The recognition results are summarized and plotted as shown in Figure 6. In these graphs, the vertical axis shows

the degree of match (DoM), while the horizontal axis indicates the corresponding input image. The first five sample images were collected from a laboratory environment, samples (6-10) were taken from a corridor, samples (11-15) were generated in a foyer and finally, samples (16-20) were gathered in an outdoor environment. It is clearly shown that the architecture using the MFM mechanism has superior performance when compared to the traditional template matching approach. The proposed architecture produced very high degree of matches (DoMs), fluctuating around 90%, while the traditional template matching approach obtained comparatively low DoMs, fluctuating between 50% to 80%.

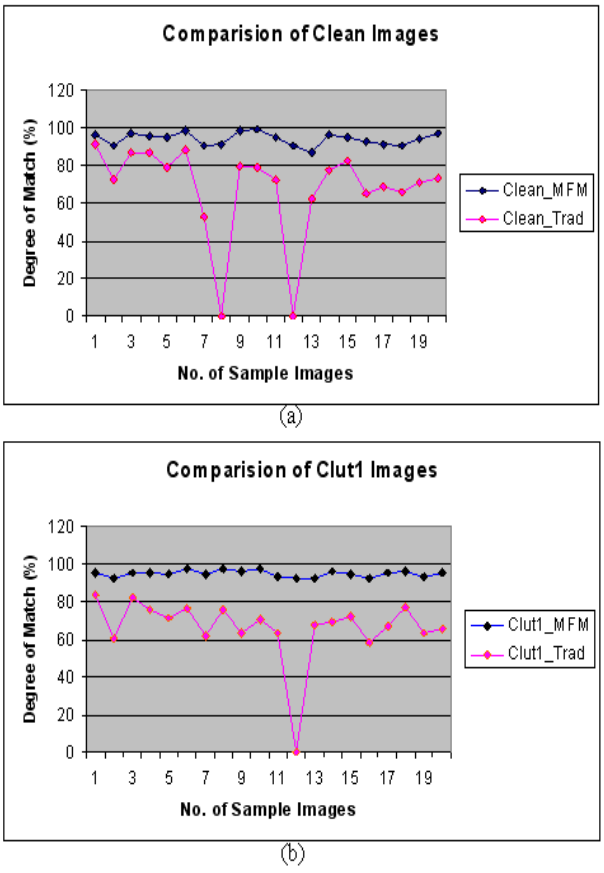

Figure 6. Comparison between the proposed landmark recognition architecture and traditional template matching approach. (a) Images of objects in clean-backgrounds. (b) Images of objects in cluttered backgrounds.

B. Partially Occluded Landmark Recognition

The proposed architecture's capability in recognising partially occluded landmark has been evaluated using a number of real images, representing different occlusion situations. These images simulate the following situations: non-occluded, single, and multiple partially occluded landmarks. In non-occluded situations, the images were taken with a target object alone in a background. For single occluded situations, the landmark was obstructed by a single object, placed either in front or to the left or right of the target. Similarly, in multiple occluded situations, multiple objects were placed in different locations (left, right and in front) to partially obscure the landmarks.

A total of twenty four images were captured and processed by the proposed architecture, and the recognition results are shown in Table.1. It records the best degree of match (DoM) value between the input ROI region and the corresponding memory template. Table.1 shows that non-occluded landmarks have very high DoM values, approximately 97% as expected, and diminishes with the increase in the level of concealment. Nevertheless, in single occlusion situations, the proposed architecture was able to maintain DoMs above 90% for all of the tested images. However, in the case of multiple occlusions, the architecture is only able to obtain DoM values over the threshold in a few cases. Future research will focus on enhancing the architecture to address the issue of recognising landmarks that are partially occluded by many objects.

Table.1: The results of the template matching stage – the match between the best ROI region in the input image and the memory template

Objects	No Occlusion(%)	Occlusion Right (%)	Occlusion Left (%)	Occlusion Centre (%)	Occlusion L&R (%)	Occlusion R ,L &C (%)
Helicopter	97.3	94.3	96.1	90.1	91.7	76.9
Clock	97.1	95.7	91.1	95	77.0	73.5
Boat	96.7	92.2	91.4	90.1	84.0	74.5
Teapot	96.0	94.7	92.3	92.5	91.5	74.5

VI. CONCLUSIONS

The paper describes a visual landmark recognition architecture that mimics properties of the human vision system using ART and SARRT artificial neural networks. It uses a set of static and dynamic binary filters to create memory-guided feedback pathways that selectively govern the feature extraction, matching, and memory activation stages. The architecture has shown superior performance over traditional template approach in recognizing occluded and non-occluded visual landmarks in both clean and cluttered backgrounds.

ACKNOWLEDGMENTS

The work described in this paper was funded by Weapons Systems Division, Defence Science and Technology Organisation (DSTO) by research contract No. 4500 177 390.

REFERENCES

[1] M. Magnard and J. G. Malpeli, "Paths of Information Flow Through Visual Cortex," *Science*, vol. 251, 1991

[2] L. Sherwood, *Human Physiology: From Cells to Systems*, Third ed: Wadsworth, 1997.

[3] R. F. Thomps, *The Brain: a neuroscience primer*, 3rd ed. New York: Worth Publishers, 2000.

[4] M. J. Tarr and S. Pinker, "When Does Human Object Recognition use a Viewer-centered Reference Frame," *Psychological Science*, vol. 1, pp. 253-256, 1990

[5] S. Grossberg, "Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, and illusions," *Biological Cybernetics*, vol. 23, pp. 187-202, 1976

[6] S. Grossberg, "How does a brain build a cognitive code?," *Psychological Review*, vol. 87, pp. 1-51, 1980

[7] S. Grossberg, "Neural expectation: Cerebellar and retinal analogs of cells fired by learnable or unlearned pattern classes," *Kybernetik*, vol. 10, pp. 49-57, 1972

[8] G. A. Carpenter and S. Grossberg, "A massively parallel architecture for a self-organizing neural pattern recognition machine," *Computer Vision, Graphics, and Image Processing*, vol. 37, pp. 54-115, 1987

[9] G. A. Carpenter and S. Grossberg, "ART 2: Self-organization of stable category recognition codes for analog input patterns.," *Applied Optics*, vol. 26, pp. 4919-4930, 1987

[10] G. A. Carpenter and S. Grossberg, "ART 3: Hierarchical search using chemical transmitters in self-organising pattern recognition architectures.," *Neural Networks*, vol. 3, pp. 129-152, 1990

[11] P. Lozo, "Neural theory and model of selective visual attention and 2D shape recognition in visual clutter," in *Department of Electrical and Electronic Engineering. Adelaide*: University of Adelaide, 1997.

[12] P. Lozo, "Selective attention adaptive resonance theory (SAART) neural network for neuro-engineering of robust ATR systems," presented at IEEE International Conference on Neural Networks, 1995, pp.2461-2466

[13] J. H. R. Maunsell and V. P. Ferrera, "Attentional Mechanisms in Visual Cortex," in *The Cognitive Neurosciences*: MIT Press, 1994, pp. 451-461.

[14] J. Moran and R. Desimone, "Selective Attention Gates Visual Processing in the Extrastriate Cortex," *Science*, vol. 229, pp. 782-784, 1985

[15] R. Desimone and J. Duncan, "Neural Mechanisms of selective visual attention," *Annual Review of Neuroscience*, vol. 18, pp. 193-222, 1995

[16] R. Desimone, "Neural mechanisms for visual memory and their role in attention," presented at Proceedings of the National Academy of Sciences, USA, 1996, pp.13494-13499

[17] R. Desimone, M. Wessinger, L. Thomas, and W. Schneider, "Attentional control of visual perception: Cortical, and subcortical mechanisms," *Cold Spring Harbour Symposium in Quantitative Biology*, vol. 55, pp. 963-971, 1990

[18] S. Ullman, *High-level Vision: Object Recognition and Visual Cognition*: MIT Press, 1996.

[19] A. M. Sillito, H. E. Jones, G. L. Gerstein, and D. C. West, "Feature-linked synchronization of Thalamic Relay Cell Firing Induced by Feedback from the Visual Cortex," *Nature*, vol. 369, pp. 479-482, 1994

[20] M. Mignard and J. G. Malpeli, "Paths of Information Flow Through Visual Cortex," *Science*, vol. 251, pp. 1249-1251, 1991

[21] J. Westmacott, P. Lozo, and L. Jain, "Distortion invariant selective attention adaptive resonance theory neural network," presented at Third International Conference on Knowledge-Based Intelligent Information Engineering Systems, IEEE Press, USA, 1999, pp.13-16

[22] Q. V. Do, P. Lozo, and L. C. Jain, "Autonomous Robot Navigation using SAART for Visual Landmark Recognition," presented at The 2nd International Conference on Artificial Intelligence in Science and Technology, Tasmania, Australia, 2005, pp.64-69