

# Colour image segmentation using the self-organizing map and adaptive resonance theory

N.C. Yeo, K.H. Lee, Y.V. Venkatesh, S.H. Ong\*

*Department of Electrical and Computer Engineering, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260, Singapore*

Received 26 March 2005; received in revised form 20 July 2005; accepted 20 July 2005

## Abstract

We propose a new competitive-learning neural network model for colour image segmentation. The model, which is based on the adaptive resonance theory (ART) of Carpenter and Grossberg and on the self-organizing map (SOM) of Kohonen, overcomes the limitations of (i) the stability–plasticity trade-offs in neural architectures that employ ART; and (ii) the lack of on-line learning property in the SOM. In order to explore the generation of a growing feature map using ART and to motivate the main contribution, we first present a preliminary experimental model, SOMART, based on Fuzzy ART. Then we propose the new model, SmART, that utilizes a novel lateral control of plasticity to resolve the stability–plasticity problem. SmART has been experimentally found to perform well in *RGB* colour space, and is believed to be more coherent than Fuzzy ART.

© 2005 Elsevier Ltd. All rights reserved.

*Keywords:* Adaptive resonance theory; Colour image segmentation; Neural networks; Lateral control; Network plasticity; Network stability; Self-organizing map

## 1. Introduction

The problem considered in this paper is the segmentation of colour images for which many methods have been proposed in the literature. Amongst the non-classical methods, the application of artificial neural networks (ANN) is prominent. In recent years, motivated by the remarkable characteristics of the human visual system (HVS), researchers have applied ANNs to various problems in pattern recognition [1]. ANNs have several advantages over many conventional computational algorithms, among which the most important are (i) massive parallelism, (ii) better adaptability to different data sets, (iii) fault-tolerance to missing, confusing and noisy data, and (iv) optimal (or ‘near optimal’) performance.

Networks for three types of classification have been employed: supervised, unsupervised and a combination of the two. In the segmentation of colour images, unsupervised learning is preferred to supervised learning because the latter requires a set of training samples, which may not

always be available. Furthermore, adaptive neural-network computing methods are more effective and efficient than traditional ones. Since the focus of this paper is on the application of competitive-learning neural networks to the problem of colour image segmentation, we review the relevant results in some detail below.

In unsupervised, self-organizing neural networks, the two dominant models are the self-organizing map (SOM) [2,3] and adaptive resonance theory (ART) [4,5], both of which are based on competitive learning.

SOM, which was originally introduced for the visual display of one- and two-dimensional data sets, has the same functional ideas as many other clustering algorithms. The SOM neural network is a topology-preserving map in which adjacent vectors in  $\mathcal{U}$  are mapped to adjacent (or identical) cells in the array, and adjacent cells in the array have similar position vectors in  $\mathcal{U}$ . The purpose of the self-organization process, as described by Kohonen [2,3], is to find values for the position vectors such that the resultant mapping is topology- and distribution-preserving<sup>1</sup>.

\* Corresponding author. Tel.: +65 68742245.

E-mail address: [eleongsh@nus.edu.sg](mailto:eleongsh@nus.edu.sg) (S.H. Ong).

<sup>1</sup> By a distribution-preserving mapping we mean, that for a random vector with the probability density function,  $p(X)$ , each cell has the same probability of being the target of the mapping. Stated otherwise this means that the relative density of position vectors in  $\mathcal{U}$  approximates the probability density of  $p(X)$ .

Dekker [6] presented the use of SOM network for quantization of colour graphics images. By adjusting a quality factor, the network is shown experimentally to produce images of much greater quality with longer running times, or slightly better quality with shorter running times than existing methods. In a refined version of the SOM, the output can be used for a controlled training of the next layer network in the manner of Lampinen and Oja [7], who proposed a multi-layer self-organizing map, HSOM, as an unsupervised clustering method. Analogous to multi-layer feed-forward networks, the HSOM (i) forms arbitrarily complex clusters, (ii) provides a natural measure for the distance of a point from a cluster by giving appropriate weights to all the points belonging to the cluster, and (iii) produces clusters that match the desired classes better than the direct SOM or the classical  $k$ -means or ISODATA algorithms.

Traven [8] has investigated the application of a competitive learning algorithm to statistical pattern classification using both local spectral and contextual features, but it is a supervised learning procedure in which the image must first be manually segmented. Ghosal and Mehrotra [9] describe a Kohonen self-organizing feature map for segmenting range images using local information provided by the orthogonal Zernike moments. However, the application of their algorithm is limited to planar and 1-D quadratic surface patches of gray level images. Papamarkos et al. [10] applied a tree clustering procedure to achieve colour reduction. In each node of the tree, a principal component analyzer and a Kohonen self-organized feature map (SOFM) neural network define the colour classes for each node. A limitation of this method is that the maximum number of final colours has to be specified *a priori*. Uchiyama and Arbib [11] employ competitive learning as a tool for colour image segmentation. After demonstrating the equivalence of vector quantization and cluster-based techniques, they apply their algorithm to gray scale and colour images. The final results appear to be essentially no different from those obtained by clustering.

A hierarchical two-stage SOM network is employed in [12] as a pattern classifier to enhance the results of conventional single-stage SOM without *a priori* information on the appropriate number of clusters to be used in the segmented image. However, the map size for both stages needs to be heuristically determined, and it is found to be difficult to achieve optimally sized maps.

The ART architecture is also a self-organizing network that allows the system to switch between a *learning* or *plastic* state (in which the network parameters may be modified) and a *stable* or *fixed* state for operation. Fig. 1 presents the general structure of the families of ART models.

The ART-based network involves three groups of neurons: an input or comparison stage, an output or recognition stage, and a mechanism to control the degree of similarity of patterns placed on the same cluster (a reset

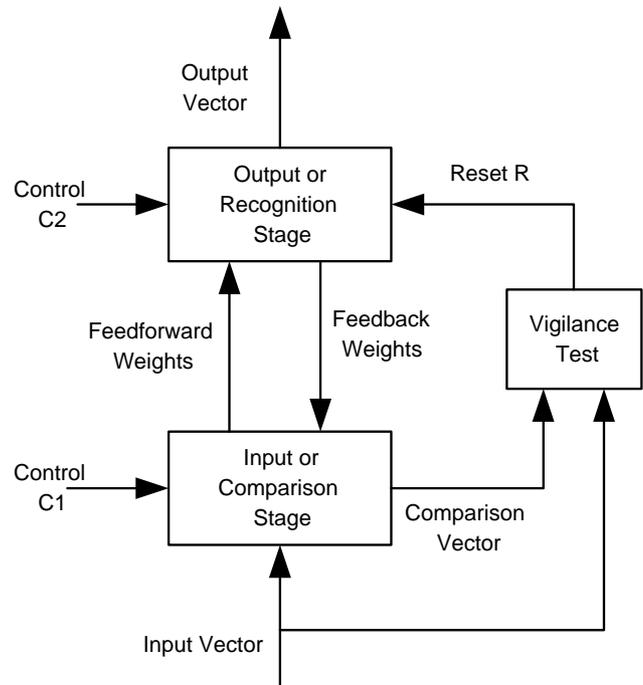


Fig. 1. General architecture of the ART.

mechanism). Each neuron in the input stage is connected to all the neurons in the output stage using feedforward weights; conversely, each neuron in the output stage is connected to all the neurons in the input stage using feedback weights. Control signals C1 and C2 along with a reset signal R facilitate comparison of the inputs with a ‘vigilance pattern’ in order to determine whether a new class pattern should be created for any given input pattern.

There are many versions of the ART model [4,5], among which we cite ART1, ART2 [13], and Fuzzy ART [14]. The first can stably learn to categorize binary input patterns presented in an arbitrary order. The second, ART2, discovers input data clusters of either analog or binary patterns (presented in an arbitrary order) without considering their actual size. It has the ability to produce hierarchical clustering that is insensitive to non-uniform variations in the input data distribution [15]. The third model, Fuzzy ART, incorporates computations from fuzzy set theory into ART1. For a detailed operation of the ART-based networks, the reader may refer to [15,16]. It should be noted here that ART does not need any pre-specified number of clusters.

ART networks are designed to be both stable and plastic, i.e. they learn a new pattern equally well at any stage of learning. The core issue in the application of ART networks, for instance, to colour image segmentation is the stability–plasticity dilemma which can be described as follows. It is desirable that the closer the network is to its converged state, the more strongly it should resist the erasing of the information learned earlier. On the other hand, if the network is far from its converged state or when there are previously unseen inputs, it should be more sensitive to the learning of any new input pattern, although this learning



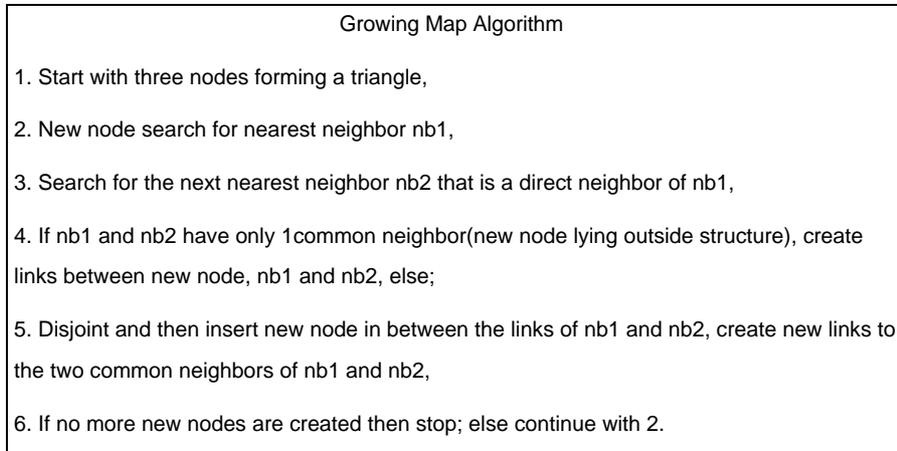


Fig. 3. The Growing map algorithm.

appropriate number of categories [17,18]. In the case of the latter (i.e. ART networks), although the number of classification categories is adaptive, (i) the ART2 network cannot reach the stage of absolute stable equilibrium and, under certain external conditions, the network can become unstable [16,19]; and (ii) the Fuzzy ART network, which uses unidirectional learning to prevent instability, is prone to statistical inconsistencies during classification [19].

Therefore, in an attempt to overcome the deficiencies of the individual systems, we develop new strategies by incorporating SOM into (Fuzzy) ART, and call the model SOMART. More specifically, we attempt to combine the key principles of SOM and ART to form a neural network that has the desirable characteristics of both, i.e. obtaining a useful feature map, and improving performance by utilizing lateral control of plasticity. The characteristics are summarized below.

SOM: (i) Feature map for the visualization of abstraction. (ii) Lateral control of plasticity.

ART: (i) Real-time (on-line) learning. (ii) Functions in a non-stationary environment. (iii) Adaptive number of clusters/map size. (iv) Plasticity in an unexpected environment. (v) Self-regulating hypothesis testing to globally reorganize the energy landscape. (vi) Fast adaptive search for best match. (vii) Rapid direct access to codes of familiar events. (viii) Adjustable discriminative ability. (ix) Scalability of the properties to arbitrarily large system capacities.

The issues that need to be addressed in order to accomplish this objective are the lateral control mechanisms and the generation of topological relations between prototypes. These are presented in Sections 2.1 and 2.2.

### 2.1. Lateral control mechanisms

For the modeling of the physiological SOM process, Kohonen [2] defined two types of lateral control:

- 1 The lateral activity control (WTA function), usually called the ‘Mexican hat’ function, or On-Center-Off-Surround contrast enhancement; and
- 2 The lateral plasticity control (neighborhood function), which defines how local activity determines the learning rate in its neighborhood. This kernel is non-negative and may take on the Gaussian form.

We shall highlight these two types of lateral control separately in the contexts of SOM and ART.

The WTA function is implemented in the SOM and ART neural networks by lateral-feedback circuits, as has been traditionally used in neural networks [2,5]. In both networks, the lateral feedback can be made via interneurons (STM) in which, for stable convergence, the time constants must be significantly smaller than those of the principal cells (LTM) in order that the interneurons converge faster than the principal cells. However, the two networks differ in the way they reset their activity while responding to new inputs. In the SOM, resetting is done automatically and locally by slow inhibitory interneurons. This WTA circuit operates in cycles, where each cycle can be thought to correspond to one discrete-time phase of the SOM algorithm. Normally the input would be changed at each new cycle. In contrast, ART uses an attention-orienting system which consists of complex interactions of STM and LTM to implement a parallel control of all neurons of the network. The system enables ART to code adaptively in response to a series of inputs without the need to present each of them at each new cycle.

In the lateral plasticity control of SOM, the ‘winner’ directly modulates the synaptic plasticity in the lateral direction. ART networks have no explicit means of lateral

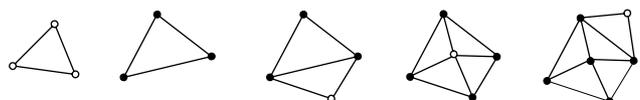


Fig. 4. Development of a cell structure.

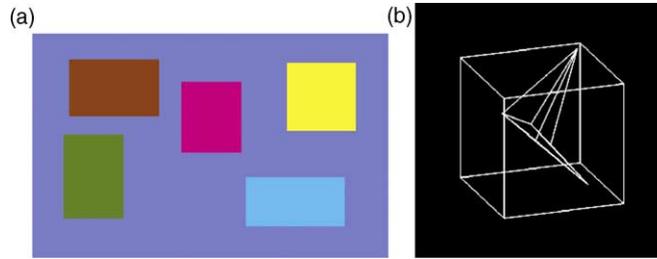


Plate 1. (a) Original ‘geometric data’ image and, (b) its topological map in geometric RGB colour space with six vertices representing the initial cluster centers for Gaussian distribution data.

plasticity control but implement plasticity in another form without direct lateral control of the ‘neighborhood’ in the F2 layer, i.e. the ability to create new categories for unfamiliar patterns. This provides for the possibility of having a lateral plasticity control scheme for ART, and also ensures that there is no conflict of lateral control mechanisms.

2.2. SOMART-generation of lateral connections using Fuzzy ART

To incorporate a topological map in ART, we build an experimental ART-based model to generate a growing topological map and call the model SOMART. The important criterion is that map lattices should be relevant to the global energy landscape. The lattice vertices should preferably be centroidal; and non-centroid vertices can be recomputed to obtain centroidal vertices. For a growing map to form coherent

topological relations, newly committed prototypes should be positioned within the pattern space they represent and not shifted excessively. Fuzzy ART with fast-learn slow-recode option is chosen over ART2 for the sake of simplicity and speed. Fuzzy ART with monotonically decreasing top-down and bottom-up weights has the limitation that the formed lattice vertices will not be centroidal but they should at least be topologically coherent [20].

ART starts with a minimum number of prototypes and adaptively increases the number of prototypes to represent the inputs. In contrast with the method proposed by Fritzke [20,21] to construct two-dimensional cell structures that are specially adapted to the underlying distribution, the Fuzzy ART algorithm, in our model, is used to determine when and where to insert cells in the current structure. A new cell is

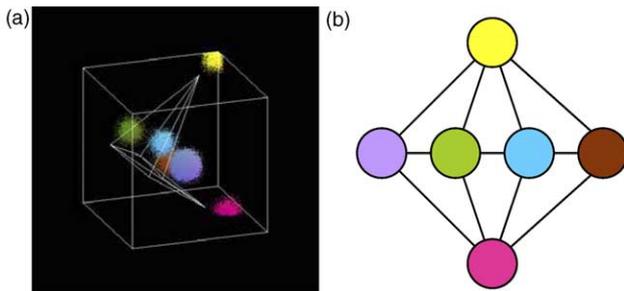


Plate 2. SOMART topological map generated using ‘geometric data’ image with Gaussian noise of standard deviation 0.001 added: (a) topological map in geometric RGB colour space with six vertices; and (b) its ‘unfolded’ 2D topological map.

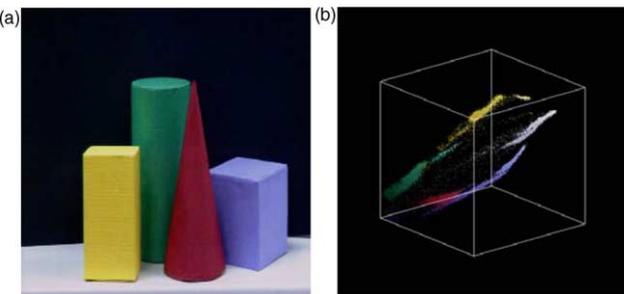


Plate 3. (a) Original ‘objects’ image; and (b) its representation in geometric RGB colour space.

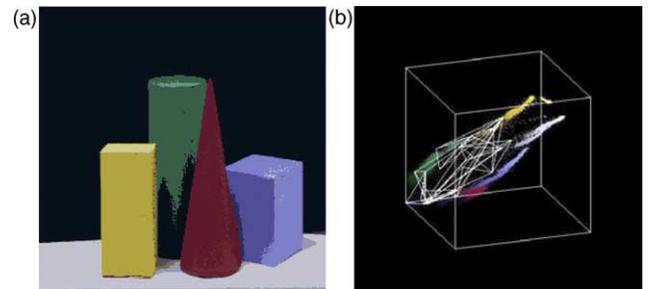


Plate 4. SOMART segmentation results on ‘objects’ image using vigilance parameter  $\rho=0.70$  with 23 clusters formed ( $\beta=0.5, \alpha=0.1$ ): (a) segmented image; and (b) its topological map in geometric RGB colour space with 23 vertices.

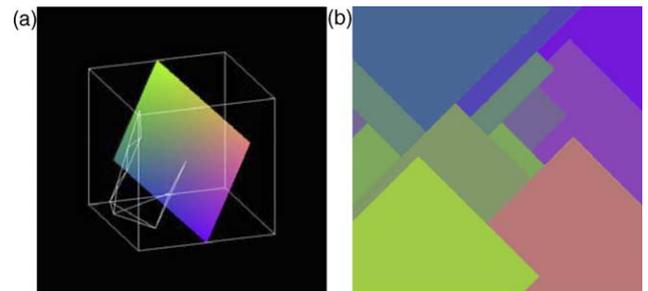


Plate 5. SOMART segmentation results on ‘continuous spanning data’ image using vigilance parameter  $\rho=0.70$  with 13 clusters formed ( $\beta=0.5, \alpha=0.1$ ): (a) topological map in geometric RGB colour space with 13 vertices; and (b) the segmented image.

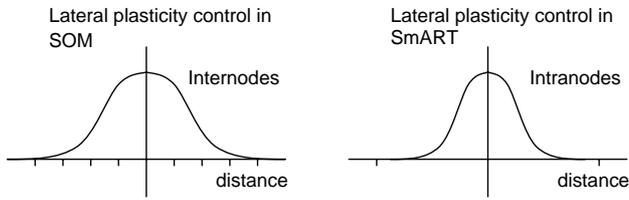


Fig. 5. The two types of lateral control of plasticity for SOM and SmART respectively: (a) internodes lateral plasticity control, (b) intranodes lateral plasticity control.

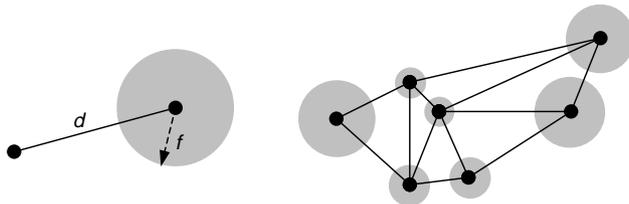


Fig. 6. The implementation of plasticity parameter in SmART: (a) degree of freedom shown as gray region (sphere), (b) feature map illustrating SmART's lateral control of plasticity.

always inserted in where the unmatched input pattern lies. This is achieved by using fast-learn in Fuzzy ART for the new cell. In conjunction with the Weber Law that the existing prototypes conserve weight values whenever possible in the conservative limit [14], the integrity of the map topographical relationships should be preserved.

The SOMART growing map algorithm and its illustration are shown in Figs. 3 and 4, respectively. With complement coding, only prototypes of the original input set

are regarded as nodes in the map. The global ordering of the feature map in SOMART is achieved by the self-regulating hypothesis testing of Fuzzy ART.

Since ART is designed to function in a real-time environment, there is no differentiation between the training and classification phases. The algorithm gives classification results simultaneously with learning. Hence, in our simulations, the classification mode is not applied. We derive and display the results in real-time as the network is being trained without changing any of its parameters until the learning process has converged.

### 2.3. Experimental results with SOMART

We analyze and evaluate the performance of the SOMART algorithm by using a 3D visual simulator. Unlike SOM, the SOMART topological map needs no initialization to form the global map. Global ordering is ensured by the underlying Fuzzy ART algorithm. The original images used before the addition of Gaussian noise and its initial cluster centers in geometric colour space are shown in Plate 1. Added Gaussian noise (standard deviation of 0.001) makes the cluster centers more visible. The geometric data set consists of 64,000 (320 × 200) patterns generated in each of six predefined kernels. Simulations show that a vigilance parameter of 0.7 is able to separate the input vectors into six clusters.

Plate 2(a) shows a topological map generated using the above image as inputs to SOMART with a learning rate of 0.5. Using topological relations, we ‘unfold’ the map to illustrate its 2D property (Plate 2(b)). It is observed that the SOMART algorithm correctly forms the 2D topological

**SmART Algorithm**

1. Start with zero cluster prototype vectors: the set  $P$  of prototype vectors is  $\{\}$ .
2. Let  $I =$  next input vector.
3. Find the closest cluster prototype vector (if any) and call this cluster vector  $T_i$ . In other words, find  $i \in P$  to minimize the Euclidean distance
 
$$d(T_i, I) = \sqrt{T_i \cdot T_i + I \cdot I + 2T_i \cdot I}$$
4. If distance threshold  $\rho < d(T_i, I)$  or if  $P = \{\}$  (no cluster prototype vectors yet), then create new cluster,  $j$ , and set  $T_j = I$ . Degree of freedom,  $f$ , for new and direct neighbor prototype vectors are updated. Set  $P = P \cup \{j\}$ . Output  $j$ . Then go to step 2.
5. Otherwise ( $T_i$  is close enough to  $I$ ), if rate of change/degree of freedom,  $f > \lambda$ ,
 

update  $T_i : (1-\lambda)T_i + \lambda I$

else  $T_i : (1-f)T_i + fI$

Output  $i$ . Go to step 2.

Fig. 7. The SmART algorithm.

Table 1  
Labeling for the different parameter ranges used in SmART

Label	Plasticity parameter	Learning rate
Low	0.01–0.30	0.001–0.010
Medium	0.31–0.80	0.011–0.100
High	0.81–1.00	0.101–1.000

map. However, when a more complex image such as the ‘Objects’ image shown in Plate 3 is used, the generated map (Plate 4(b)) forms too many irrelevant links.

It is observed that as some prototype vectors move towards the origin, other prototype vectors can take on the values that these prototypes previously had. From this characteristic, we infer that the overlapping of the prototypes’ response regions inherent in Fuzzy ART [14] causes this phenomenon. Thus relevant neighborhood links that are formed initially may be rendered irrelevant after some learning cycles but must be kept intact to preserve the property of the 2D map. This adds undue complexity to the topological map. Moreover, the prototype vectors are not representative of the cluster-means as shown in Plate 5(a), and if these cluster-means are needed, extra computational effort will be necessary to calculate them.

The statistical inconsistency of Fuzzy ART and its non-centroid clusters affect the coherence of the topological map in SOMART. With noisy data, the prototypes, being minima, tend to degenerate to zero, and new clusters must be continually created, thus resulting in category proliferation which Carpenter et al. [14] deal with by complement coding in which each input set is doubled in length. Fuzzy ART employs hyperbox clustering approach, thereby leading to the segmentation result shown in Plate 5(b).

To summarize, our objectives for incorporating SOM into ART have only been partially fulfilled. Although a growing 2D topological map using Fuzzy ART has been observed, we have not utilized the topological relations (in the bid to improve ART-based network performance) due to their inconsistent characteristics.

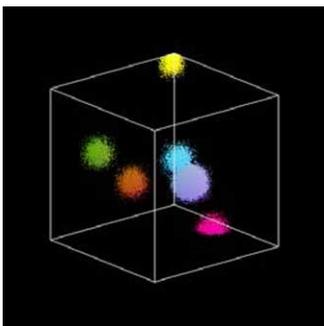


Plate 6. Gaussian data set (standard deviation 0.001) in geometric RGB colour space.

### 3. SmART– ART using euclidean metric

The drawbacks of Fuzzy ART and ART2 can be avoided if (i) the basic ART1 architecture can be retained and yet remain stable for analog inputs; and (ii) the Euclidean distance metric is employed for the calculations of nearest prototypes and thresholds, thereby reducing the possibility of categories overlapping each other.

Further, in an attempt to resolve the two issues of the plasticity–stability dilemma and the lateral control of plasticity, we propose the integration of SOM principles with ART. This leads us to our model, SmART, which utilizes a novel form of lateral plasticity control.

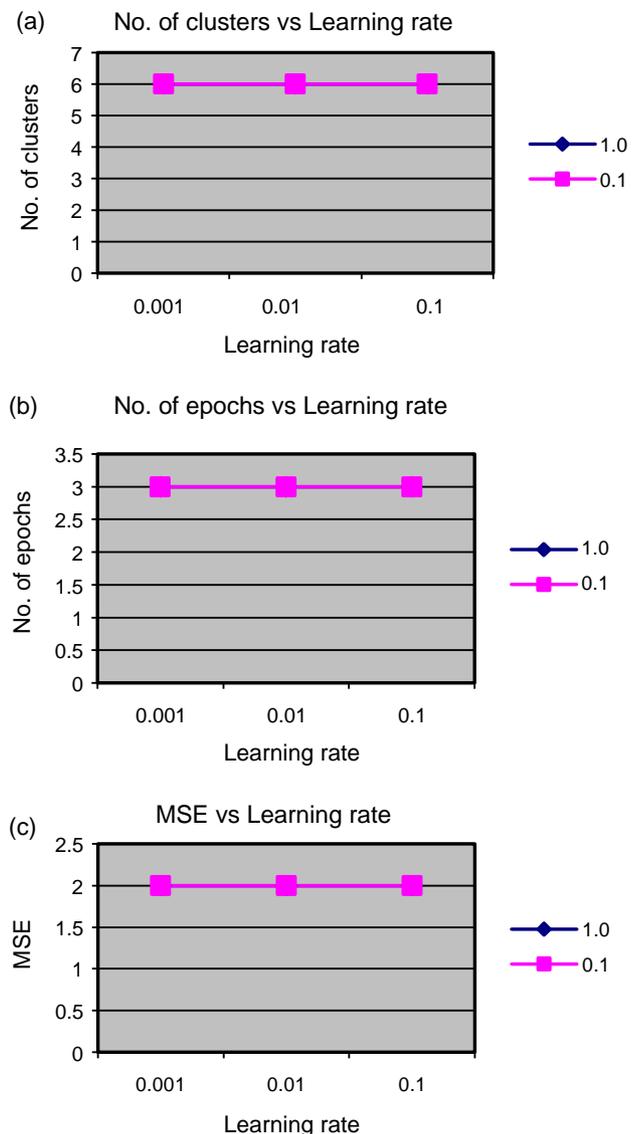


Fig. 8. SmART segmentation results on ‘geometric data’ image with Gaussian noise of standard deviation 0.001 added: (a) No. of clusters formed vs. different learning rates; (b) No. of epochs needed for convergence vs. different learning rates; (c) MSE vs. different learning rates. Results are shown at two different plasticity values of 0.1 and 1.0. Vigilance is set at 0.36 and standard stopping criterion at first zero ART reset is applied.

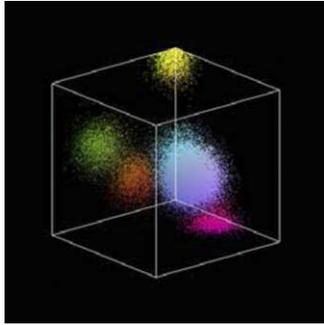


Plate 7. Gaussian data set (standard deviation 0.002) in geometric RGB colour space.

We overcome the problem of instability in ART1, when Euclidean distance is used (see [22] for definitions of stability and instability), by utilizing the topological map developed for SOMART, and implementing the lateral control of plasticity in Euclidean ART1, without resorting to the highly elaborate approach found in ART2.

### 3.1. Lateral control of plasticity (plasticity parameter)

Instead of using the topological map to implement lateral plasticity control (neighborhood function) as in SOM, the topological relations between nodes are used as an adaptive learning inhibitory function on the prototype vectors. The difference between the lateral control of plasticity in SOM and SmART is illustrated in Fig. 5. This approach prevents overmatching or the overwriting of the response region of WTA onto the regions of its topological neighbors, i.e. we preserve previously stored information when a new pattern is being learned. The function of the topological map in SmART, besides providing a feature map, is to enable the lateral control of plasticity in a manner different from that found in SOM. Most important of all, the condition on ART that only the WTA unit learns during resonance is not violated here.

We model the plasticity–stability dilemma with a single plasticity parameter that governs the lateral control of plasticity of the network. The plasticity parameter utilizes the topological map to provide the distance measures

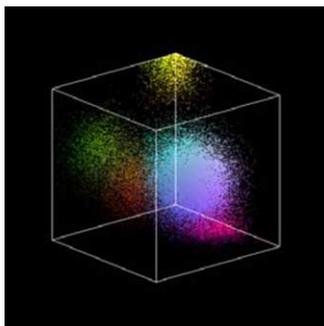


Plate 8. Gaussian data set (standard deviation 0.003) in geometric RGB colour space.

Table 2

Suggested values of parameters for SmART with specified vigilance on images of varying complexity

Colour cluster feature	Plasticity parameter	Learning rate
Well-separable	Minimal effect	Minimal effect
Partially-separable	High	Low
Non-separable	Medium	Low

between neighboring prototypes at the classification layer of ART. If  $f$  is the degree of freedom for a node,  $I$  the plasticity parameter and  $d$  the Euclidean distance of the node’s shortest direct neighborhood link, then  $f = \delta \times d$ .

The degree of freedom for each and every node, computed from the plasticity parameter and the distance from their shortest direct neighborhood, is incorporated into the topological map to prevent oscillations, proliferation of prototypes, and, implicitly, the creation of unqualified topological relations between prototypes. By calculating the maximum allowable degree of freedom for each epoch at a specified rate of change, we make the simplifying assumption that a prototype will be bounded by the degree of freedom for the epoch. Thus prototypes are held in a mesh that can evolve over time but avoids spurious changes (Fig. 6).

### 3.2. SmART algorithm

We implement the SmART algorithm (Fig. 7) by modifying the visual simulator developed for SOMART (Section 2.2). SmART replaces Fuzzy ART as the underlying algorithm to generate a topological map. We adopt Moore’s method of incorporating Euclidean distance into ART1 [22] in our SmART algorithm in which the plasticity parameter determines the strength of lateral plasticity control. For a fixed pattern set, the degrees of freedom for all prototypes are computed at the beginning of every epoch and remain fixed throughout that epoch unless new prototypes are created. The degree of freedom for the first prototype must always be zero as there is no shortest direct neighborhood link, i.e. the distance is zero.

In order to evaluate stability in the SmART algorithm, we need to devise a suitable stopping criterion. Two measures may be used to judge the state of convergence of the network. One is to examine the weight changes of the prototype vectors, but, in view of the plasticity property of SmART, this cannot be done. The other method is to use the number of ART resets. The number of resets correlates closely with the state of convergence of an ART algorithm. A novel input will cause a reset whenever a winning prototype vector fails to encode it. Thus the number of resets is indicative of the network convergence on a pattern set. By making use of this property, we choose a stopping criterion that is associated with the number of resets for each training epoch. The stopping criterion that we adopt for SmART is

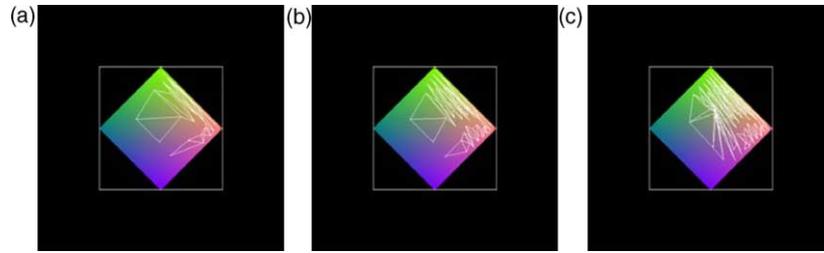


Plate 9. Snapshots of topological map in geometric RGB colour space of SmART training with high plasticity parameter on ‘continuous spanning data’ image at various training epochs ( $\rho=0.3$ ,  $\beta=0.001$ ,  $\delta=1.0$ ): (a) at 10th training epochs with 31 vertices; (b) at 20th training epochs with 61 vertices; (c) at 30th training epochs.

simply that the number of ART-resets hits zero for a training epoch.

### 3.3. Results and comparisons

Besides vigilance and learning parameters, SmART employs a plasticity parameter which is defined by heuristics rather than in any statistically meaningful way. The SmART algorithm is tested on a variety of images: Gaussian distribution data, continuous-spanning data, ‘Girl’, ‘Lena’, ‘Sails’, ‘Objects’ and ‘Fruits’. The first two synthetic images are used to evaluate the SmART algorithm. The continuous-spanning image is utilized to underscore the algorithm stability and to make a brief comparison with the SOMART algorithm. The natural images are used to analyze and characterize SmART clustering performance.

Results are presented where appropriate for each image data set as a segmented image, a graphical 2D topological map in 3D geometric colour space and in terms of a distortion (quantization, reconstruction) error measure, identified as the mean square error (MSE). See Table 1 for the heuristic labeling of the range of parameters.

#### 3.3.1. Gaussian distribution data

Fig. 13 shows the original image used for the Gaussian distribution data and the initial cluster centers in geometric colour space. The geometric data set consists of 64,000 ( $320 \times 200$ ) patterns generated in each of six predefined kernels. Simulations show that a vigilance parameter of 0.36 separates the input vectors into six clusters.

3.3.1.1. Gaussian noise with standard deviation of 0.001 added (well-separable). Plate 6 shows the same image with added Gaussian noise of standard deviation 0.001. The clusters are well-separated with the location of kernels being clearly visible. The simulation results of Fig. 8 show that, when the input vectors are well separated, different plasticity parameters and learning rates have minimal effect on clustering results.

3.3.1.2. Gaussian noise with standard deviation of 0.002 added (partially-separable). Plate 7 shows a Gaussian

distribution image with a standard deviation of 0.002. The clusters are still separable with the location of kernels remaining discernible. Generally, as input vector variance becomes larger but still remains separable, more prototype vectors are formed. It is also observed that MSE increases when compared with the previous, well separated case.

The reduction of the plasticity parameter to 0.1 has the effect of moderating the different learning rate to achieve the results at a low learning rate. As observed, the three aspects of performance are quite consistent at different learning rates. At a high learning rate, the use of high plasticity is observed to have an adverse effect because more prototypes than necessary are created. At high plasticity, as learning rate is increased, more prototype vectors are formed and the number of training epochs increases. Due to

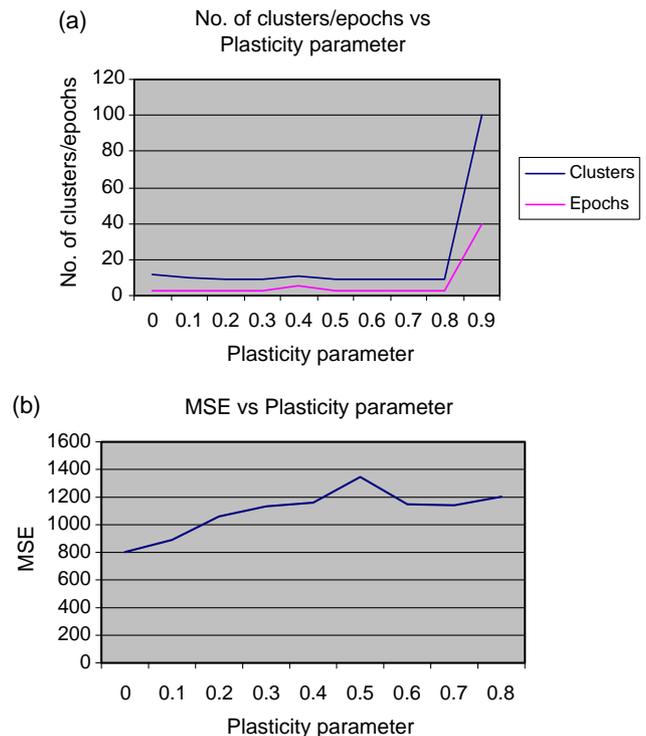


Fig. 9. SmART segmentation results on ‘continuous spanning data’ image: (a) Number of clusters/epochs vs. plasticity parameter; (b) MSE vs. plasticity parameter. Vigilance is set to 0.3 and low learning rate (0.001) is applied.

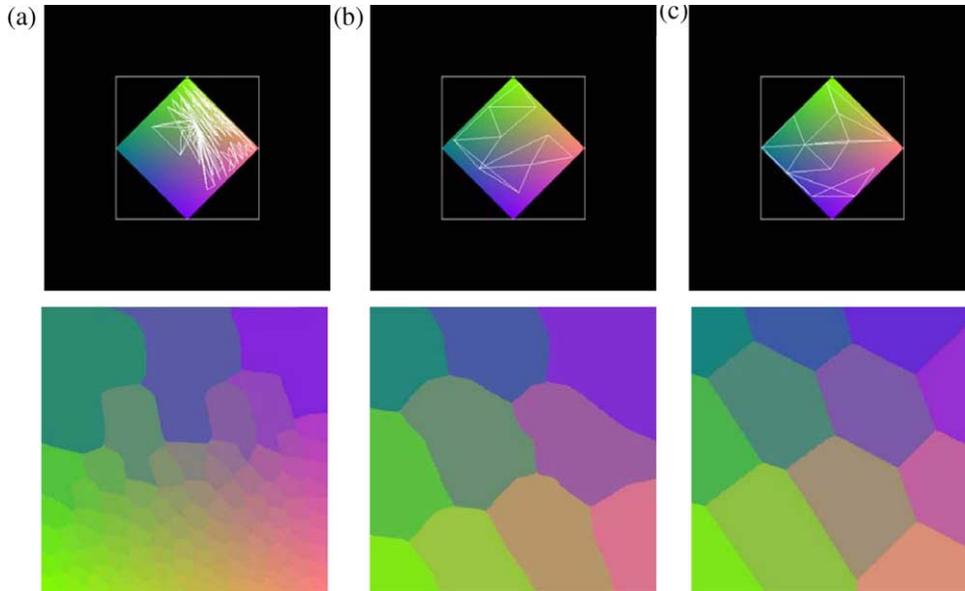


Plate 10. SmART segmentation results on ‘continuous spanning data’ image at various plasticity parameters shown as topological maps in geometric RGB colour space on the left and segmented images on the right ( $\rho=0.3, \beta=0.001$ ). The respective plasticity parameters are: (a) 1.0 with 103 vertices/clusters (termination at 35th training epoch); (b) 0.1 with 10 vertices/clusters (convergence after two training epochs); and (c) 0.01 with 12 vertices/clusters (convergence after two training epochs).

the high plasticity of these prototypes, they learn the most recent inputs very effectively, and thus frequently have the tendency to drift away from the centroids of the clusters they represent. Since, these prototypes are not representative of the cluster means, new inputs already belonging to these clusters are more likely to trigger the creation of new prototypes. In terms of MSE, the clustering performance deteriorated in spite of the increase in the number of prototypes.

We suggest the use of a low learning rate for most static applications. Since, newly committed prototypes fully converge to equilibrium values, only incremental adjustments are needed to adapt or represent clusters better as more inputs are presented. Hence the suggested approach is to use a low learning rate and high plasticity for input patterns of similar nature.

*3.3.1.3. Gaussian noise with standard deviation of 0.003 added (non-separable).* Plate 8 shows a Gaussian distribution image with a standard deviation of 0.003. The clusters overlap one another and are not clearly separable with the locations of the kernels somewhat Fuzzy. Most of the observations in the previous subsection remain valid for this non-separable case. Simulation results show that MSE is reduced across the board with fewer prototype vectors being formed when a medium plasticity parameter is used. Again, a low learning rate provides the best clustering performance that validates our previous proposal.

Hence the suggested approach here is to use a low learning rate with medium plasticity for complex non-separable input patterns of similar nature, i.e. images with

significant colour variation. A summary of the results on Gaussian distribution data is presented Table 2.

*3.3.2. Continuous-spanning data*

To evaluate stability for different plasticity parameters, we choose an image that can test the stability boundary of SmART algorithm. The continuous-spanning data of Plate 9 demonstrates that the proliferation of SmART prototype vectors can occur when high values of plasticity are used. The input vectors are generated in such a way that each prototype vector can be brought arbitrarily close to one another, thereby defeating the purpose of having a vigilance parameter. However, it should be noted that the algorithm will always converge fully with plasticity value of less than 1.0 for any input patterns.

To overcome this problem, we gradually reduce the plasticity parameter as shown in Fig. 9(a). Once the plasticity parameter is reduced to less than 0.8, the proliferation of prototype vectors is observed to be contained. As implicitly shown in Fig. 9(b) for images with large portions of continuously varying colour, a lower plasticity parameter is more likely to achieve better results. The feature of the lateral plasticity parameter to effectively control plasticity-stability balance is demonstrated here.

Table 3  
Suggested values of parameters for SmART with specified vigilance on image of specified complexity

Colour cluster feature	Plasticity parameter	Learning rate
Continuous-spanning	Low	Low



Plate 11. Natural test images. (a) 'Girl'; (b) 'Lena'; (c) 'Sails'.

Next, we assess properties of the segmented images shown in [Plate 10](#) when different plasticity parameters are applied. The SmART algorithm coherently segments the image into different partitions of dominant colours. When a lower plasticity value is employed, the order of presentation affects the classification results to a smaller extent. It can be seen that the region boundaries are more rigidly defined, and there is almost no overlapping of response regions. The use of the plasticity parameter not only ensures stability but also makes SmART more robust to order of presentation on classification. The range of values suggested for SmART plasticity parameter and learning rate are summarized in [Table 3](#).

The advantages of SmART over SOMART are illustrated by [Plate 5](#) which shows a result of the continuous-spanning image segmented using SOMART. When compared with the result obtained with SmART, its drawbacks become apparent. Although a stable equilibrium is attained, the results are obviously not as coherent as those of SmART, and its generated feature map is also less intuitive to understand.

### 3.3.3. Natural images

Until now, our simulations have been done using only synthetic images. The purpose of this section is to characterize the SmART algorithm using natural images.

#### (a) 'Girl', 'Lena' and 'Sails' images

We apply heuristics using the 'Girl', 'Lena' and 'Sails' images ([Plate 11](#)) to suggest a valid working range for SmART parameters. To this end, we set the maximum value of 1.0 for both plasticity parameter and learning rate to derive the boundary characteristic of the vigilance threshold for all the images, as shown in [Fig. 10](#). For images 'Girl' and 'Lena', a vigilance threshold of 0.3 is observed to strike a good compromise between the number of clusters, number of training epochs, computation time and MSE. Using the same criteria, a vigilance threshold of 0.4 is selected for the 'Sails' image. With these values of vigilance, we proceed to study the effects that plasticity parameter and learning rate have on MSE and the number of clusters formed. The results are shown in [Fig. 11](#). It is seen for all three images that MSE can be greatly improved in tandem with a corresponding drop in the number of representative clusters by an appropriate tuning of the plasticity parameter and learning rate. With the values of plasticity at around 0.8 or lower, SmART displays markedly more consistency in that an increase in number of clusters corresponds to a decrease in MSE across a broad range of learning rates. It can therefore be inferred from the graphs of [Fig. 11](#) that valid plasticity values for this image range from 0 to 0.8, which is consistent with earlier observations. [Plate 12–14](#) show, respectively, the segmented results of the 'Girl', 'Lena' and 'Sails'

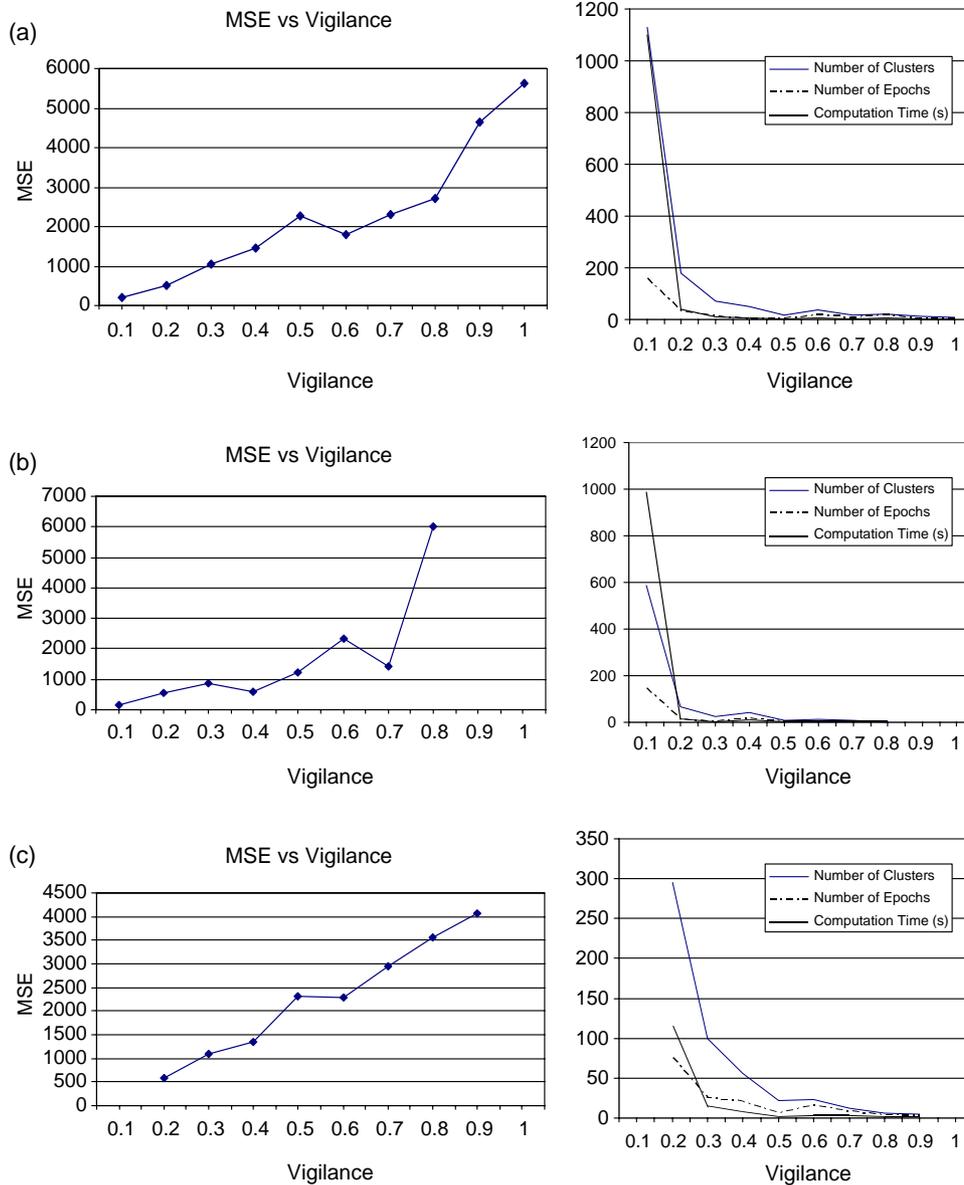


Fig. 10. SmART segmentation results illustrated with plots of MSE vs. Vigilance on left and number of clusters/epochs/computation time vs. vigilance (right) on right: (a) on 'girl' image; (b) on 'Lena' image; (c) on 'sails' image. Learning rate and plasticity parameter are all set to 1 (upper bound).

images at various learning rates with plasticity value fixed at 0.8 and vigilance threshold at 0.3.

For the sake of simplicity, the range of values suggested for the SmART plasticity parameter and learning rate is summarized in Table 4. The use of the plasticity parameter allows high learning rates to be utilized without risking instability. These simulations show that when different plasticity parameters are selected for plasticity-stability reasons, the demonstrated utility of Euclidean ART is retained.

The results also suggest that lower values of the learning rate tend to give better segmentation results in terms of MSE for a given vigilance threshold and plasticity parameter. Hence to generalize, good results are obtained with higher plasticity parameter values and

low learning rates. It is observed that with the lateral control of plasticity, instability can be effectively reduced while maintaining high plasticity. However, it appears that a non-stationary environment is needed (which is not the subject of this paper) to demonstrate the full potential of the plasticity parameter.

(b) 'Objects' and 'fruits' images

We evaluate and compare SmART performance on two images of real objects under artificial settings. We first study the effects that plasticity parameter have on the number of clusters formed at different vigilance parameters with the illustration of 2D topological maps generated in 3D geometric colour space. The SmART-segmented images are then evaluated by visual inspection to gauge their subjective segmentation quality.

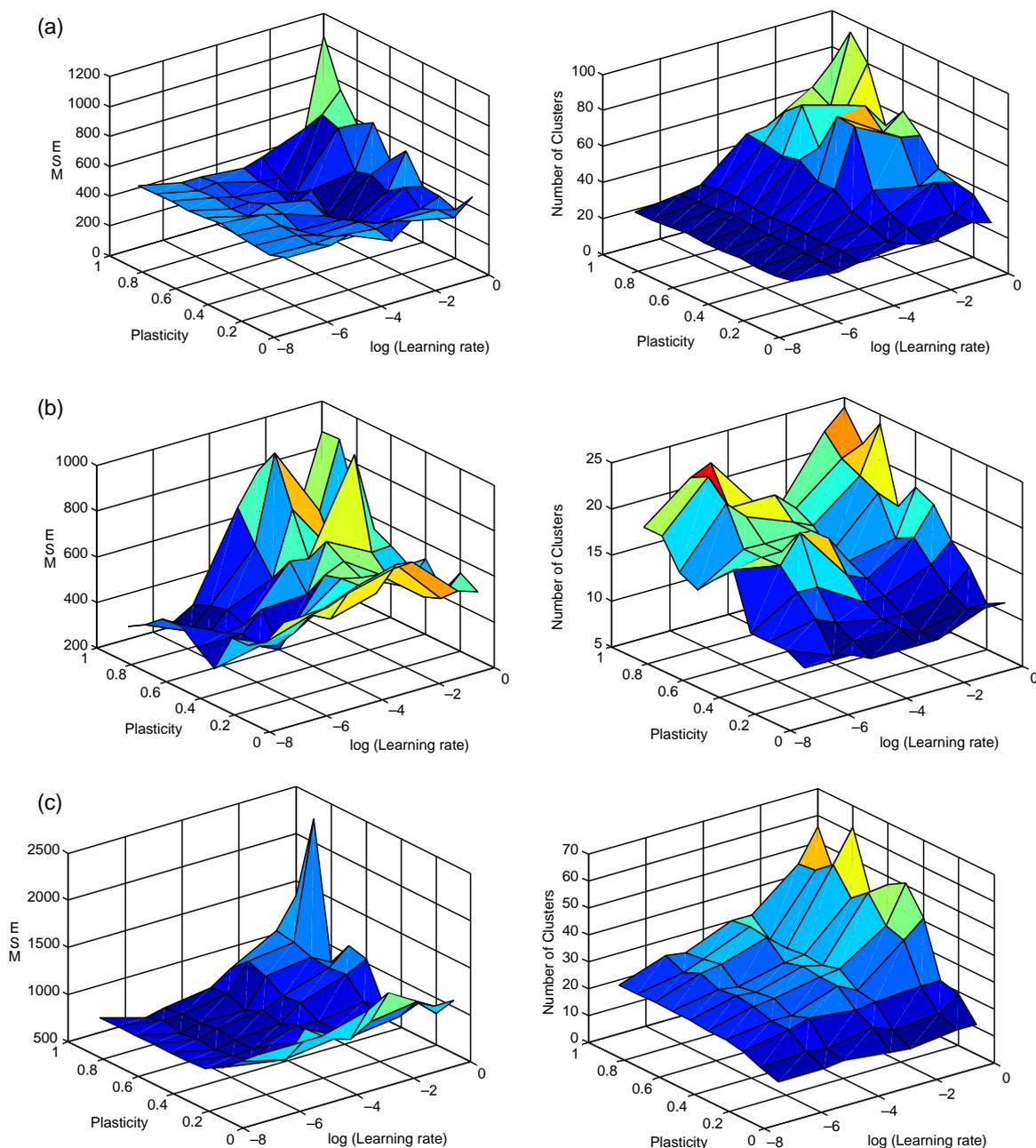


Fig. 11. SmART segmentation results illustrated with 3D plots of MSE vs. Plasticity vs. log(learning rate) on left and number of clusters vs. Plasticity vs. Log(learning rate) on right: (a) on 'girl' image; (b) on 'Lena' image; (c) on 'sails' image. Vigilance parameter is set to 0.3 except for (c) which is set to 0.4.

The 'Objects' image is of moderate complexity consisting of six dominant colours. We observe that plasticity has very little influence on the segmentation results (Fig. 12(a)). On the other hand, the 'Fruits' image is complex with 13 dominant colours and a total of 58,930 colours. With higher plasticity, SmART adaptively forms an appropriate number of clusters (Fig. 12(b)). For the more complex image, more clusters are formed at higher plasticity.

By visual inspection of the segmented 'Objects' image, the best compromise between segmentation fidelity and the smallest number of clusters is found to be at the vigilance value of 0.25 (Plate 15). Using this vigilance value at

different plasticity parameters on the same 'Objects' image (Plate 16), we find it hard to discern any meaningful differences between the segmented images.

We apply the same criteria to the 'Fruits' image to obtain the segmented image shown in Plate 17 at the vigilance value of 0.3. As before, different plasticity parameters are applied to examine their effects on the segmented images (Plate 18). It is observed that higher plasticity parameter value gives the nearest segmented representation of the original image with more clusters being formed. The results show that higher plasticity enables SmART to be more adaptive to the complexities of the input vectors.



Plate 12. SmART segmentation results on 'girl' image at various learning rate shown as segmented images ( $\rho=0.3$ ,  $\delta=0.8$ ): (a) at learning rate of 0.1 with 67 clusters/vertices (MSE=227.92); (b) at learning rate of 0.01 with 26 clusters/vertices (MSE=326.96); (c) at learning rate of 0.001 with 20 clusters/vertices (MSE=419.78).



Plate 13. SmART segmentation results on 'Lena' image at various learning rate shown as segmented images ( $\rho=0.3$ ,  $\delta=0.5$ ): (a) at learning rate of 0.1 with 12 clusters/vertices (MSE=562.29); (b) at learning rate of 0.01 with 19 clusters/vertices (MSE=259.11); (c) at learning rate of 0.001 with 17 clusters/vertices (MSE=256.97).

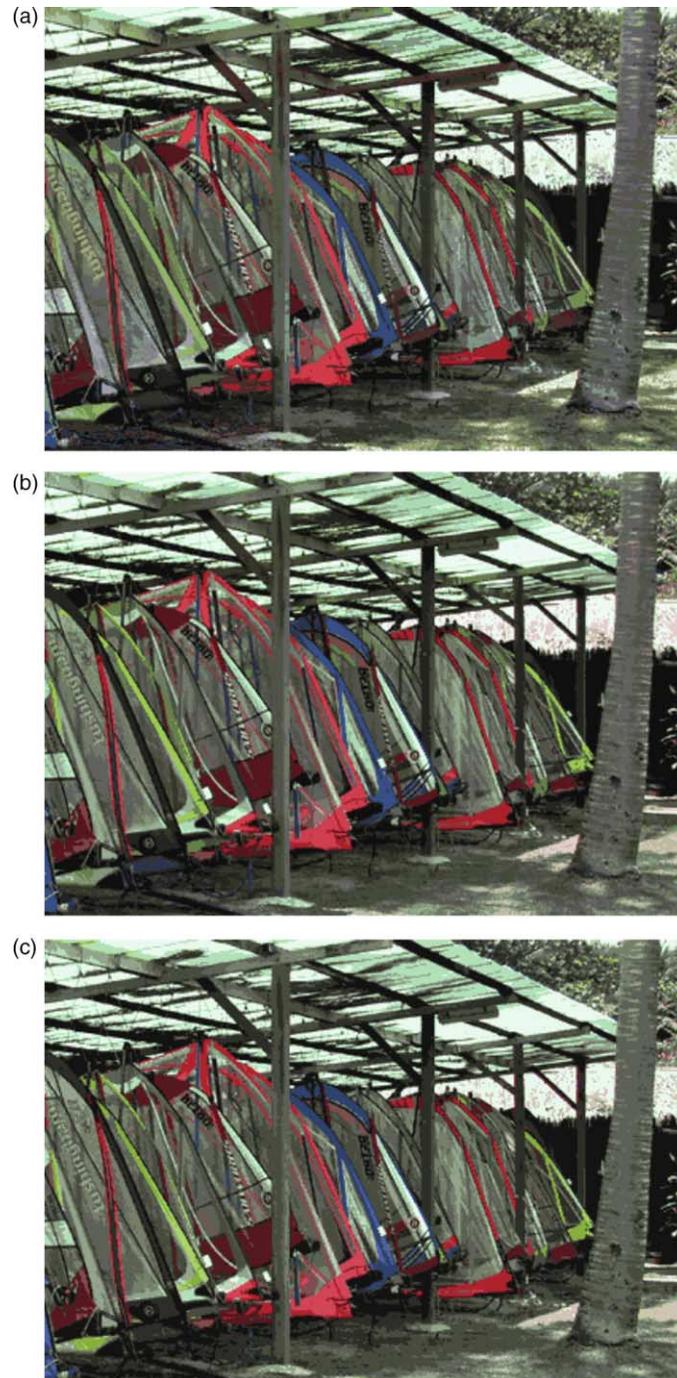


Plate 14. SmART segmentation results on ‘sails’ image at various learning rate shown as segmented images ( $\rho=0.4, \delta=1.0$ ): (a) at learning rate of 0.1 with 30 clusters/vertices (MSE=771.17); (b) at learning rate of 0.01 with 23 clusters/vertices (MSE=564.36); (c) at learning rate of 0.001 with 18 clusters/vertices (MSE=661.98).

The range of values suggested for the SmART plasticity parameter and learning rate is summarized in Table 5.

3.4. Comparison with SOMART and other architectures

In comparison with SOMART (Fuzzy ART), SmART classifications are statistically more coherent with prototype vectors representing the cluster means. Fewer prototypes are

Table 4  
Suggested values of parameters for SmART with specified vigilance on images of varying complexity

Image colour feature	Plasticity parameter	Learning rate
Complex natural images	Medium	Low

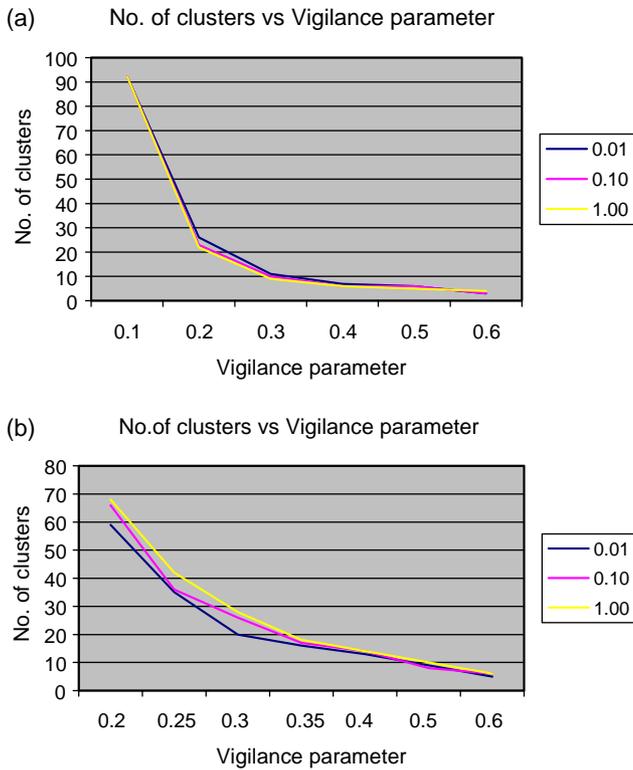


Fig. 12. SmART training showing the number of clusters vs. vigilance parameter at different plasticity parameter settings: (a) on 'object' image, (b) on 'fruits' image. Low learning rate (0.001) is applied.

needed to represent the input space since the response regions of the prototypes of SmART do not overlap one another. Therefore, SmART generates a simpler and more coherent 2D feature map that is potentially useful for visualization and abstraction purposes.

The lateral control of plasticity is implemented in SmART by utilizing the topological relationships between prototypes from the feature map, and this introduces a new plasticity parameter. Although the plasticity parameter displaces the choice parameter in SOMART, the functions that the parameters perform are entirely different. Both SmART and SOMART have three parameters while ART2 has more than eight parameters. In order to understand more intuitively the influence of the different SmART parameters, we treat (i) the vigilance parameter as defining a prototype response region, (ii) the learning rate as controlling the prototype's rate of change, and (iii) the plasticity parameter as imposing a bound on this rate of change adaptively, depending on its distance to neighboring prototypes. In other words, vigilance provides the discriminative adjustment control, learning rate determines the global adaptivity of prototypes and, with overriding priority over learning rate, the plasticity parameter restricts the adaptivity of prototypes locally, when necessary, to ensure stability. By comparison, in an ART2 network, vigilance and learning rate together with many other parameters are closely coupled together in determining stability.

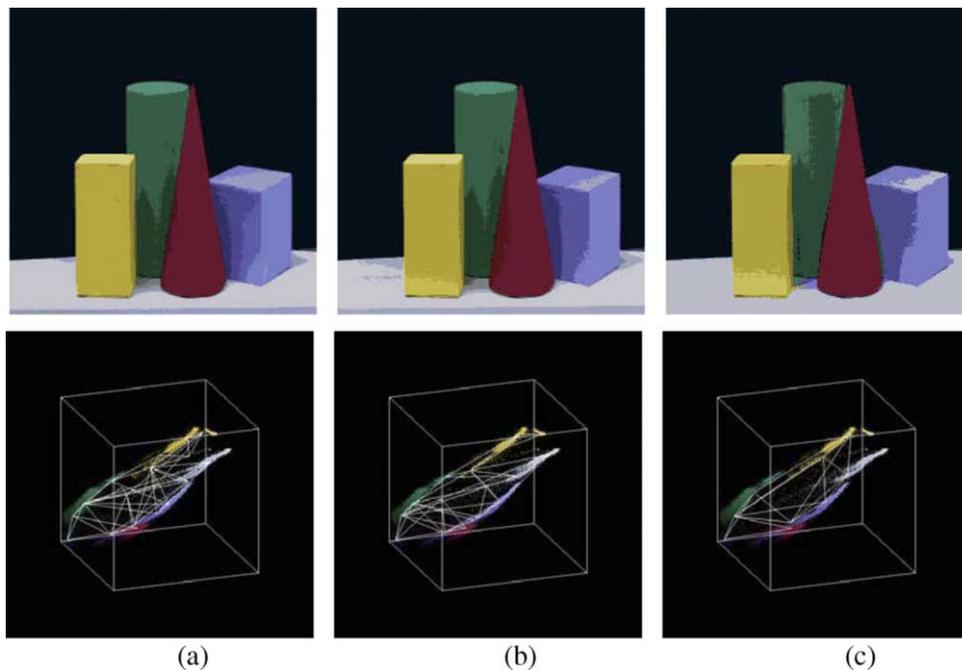


Plate 15. SmART segmentation results on 'objects' image at various vigilance parameters shown as segmented images on the left and topological maps in geometric RGB colour space on the right ( $\beta=0.001$ ,  $\delta=0.1$ ). The respective vigilance parameters are: (a) 0.2 with 23 clusters/vertices; (b) 0.25 with 18 clusters/vertices; and (c) 0.3 with 10 clusters/vertices.

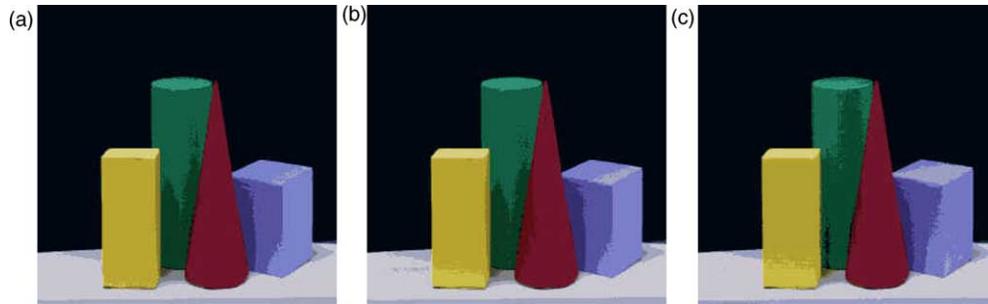


Plate 16. SmART segmentation results on 'objects' image at various plasticity parameters shown as segmented images ( $\rho=0.25$ ,  $\beta=0.001$ ). The respective plasticity parameters are: (a) 1.0 with 17 clusters; (b) 0.1 with 18 clusters; and (c) 0.01 with 18 clusters.

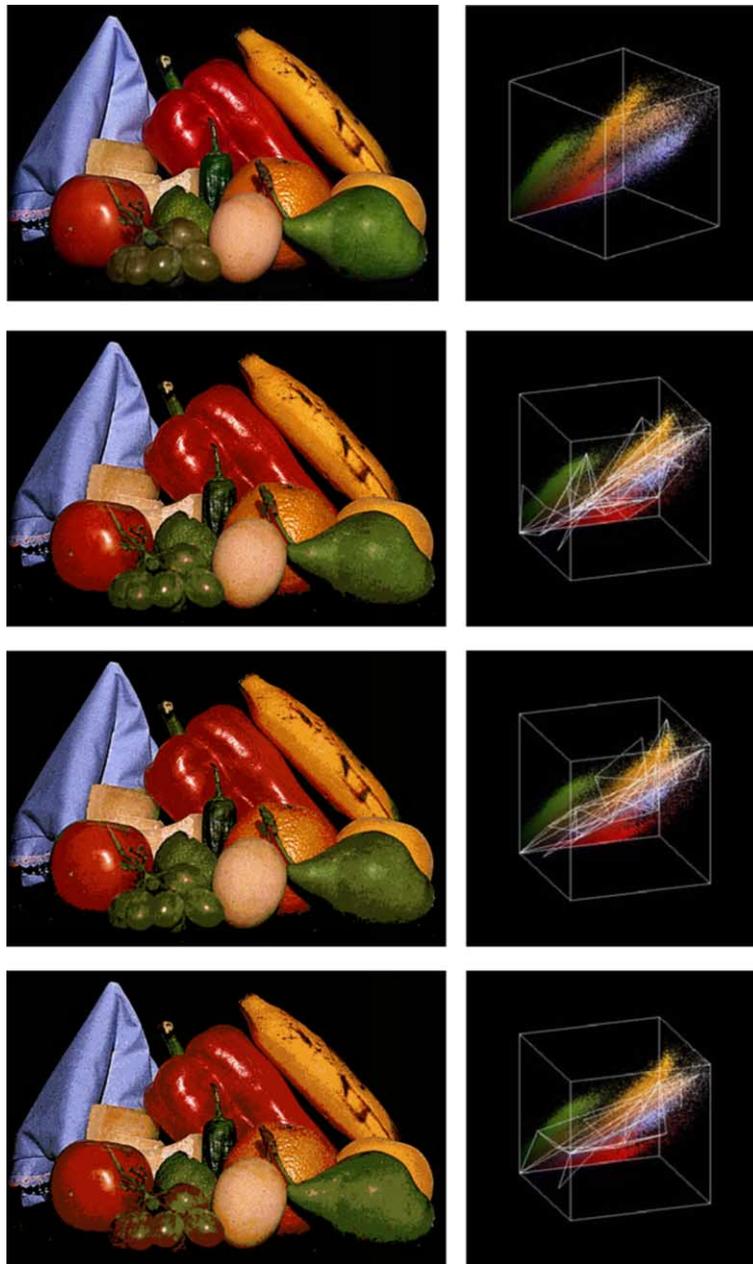


Plate 17. SmART segmentation results on 'fruits' image at various vigilance parameters shown as segmented images on the left and topological maps in geometric RGB colour space on the right ( $\beta=0.001$ ,  $\delta=0.1$ ): (a) Original 'fruits' image and its input data set in geometric RGB colour space. The respective vigilance parameters are: (b) 0.25 with 36 clusters/vertices; (c) 0.30 with 26 clusters/vertices; and (d) 0.35 with 17 clusters/vertices.

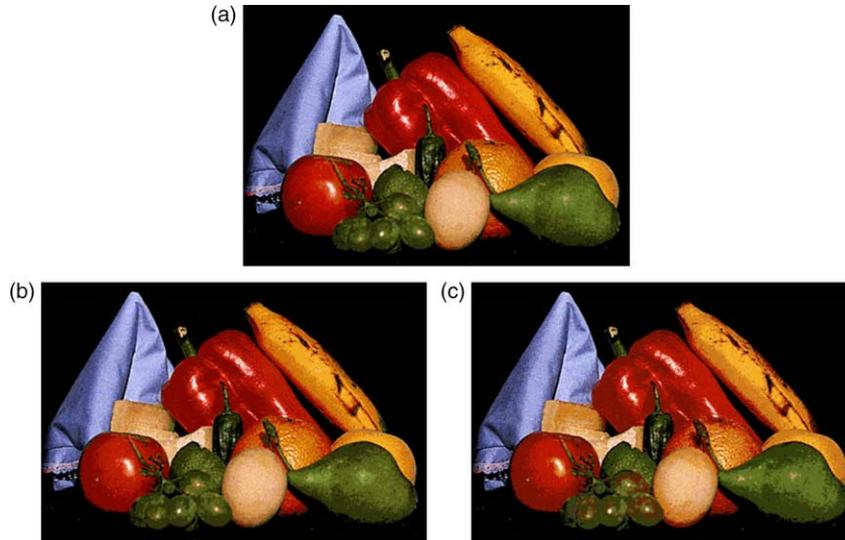


Plate 18. SmART segmentation results on ‘fruits’ image at various plasticity parameters shown as segmented images ( $\rho=0.3, \beta=0.001$ ). The respective plasticity parameters are: (a) 1.0 with 28 clusters; (b) 0.1 with 26 clusters; and (c) 0.01 with 20 clusters.

The convergence speed of SmART algorithm compares well with that of Fuzzy ART, which is in turn faster than clustering algorithms like SOM and ART2 in related colour image applications. Since, Fuzzy ART and SmART are both based upon ART1 architecture, their convergence speeds are mostly comparable except for the extra computations required for the lateral control of plasticity in SmART. It should be noted that in our implementation, the computation to determine the degree of freedom for all nodes, barring newly committed nodes and their immediate neighboring nodes, is performed only once at the end of each training epoch. This approach enhances the computational efficiency of the SmART algorithm.

The results obtained by applying SmART and ART2 to the ‘Girl’ image are shown in Table 6. ART2 parameters are adjusted by trial and error until the number of clusters approximates SmART results with the MSE minimized. ART2-training stops when the number of ART2 node-resets drops to less than 5% of the initial number of resets at the start of training. It can be seen that SmART provides a much better set of MSE results than ART2 at any learning rate with an almost equivalent number of clusters. Computational time is also comparable with ART2 without taking into consideration the advantage of a less stringent stopping criteria adopted for ART2.

The use of the plasticity parameter not only ensures stability but also enables SmART to be robust to different orders of presentation for classification. The plasticity parameter indirectly regulates the two types of plasticity: learning rate and the commitment of new nodes. It is observed that the plasticity parameter in SmART is highly effective in ensuring stability while maintaining the superior clustering performance of Euclidean ART over both Fuzzy ART and ART2.

#### 4. Conclusions

We have proposed a new neural architecture, SmART, which is based on ART1 using a Euclidean distance metric and the lateral control of plasticity, and have applied it to the problem of colour image segmentation. By utilizing the topological map developed using SOMART (which is a combination of SOM and ART), we have developed a novel method to implement the lateral control of plasticity in SmART and introduced a plasticity parameter to deal with the stability–plasticity problem.

Some salient features of SmART are: (i) minimal pre-processing of input; (ii) greater speed, stability and flexibility due to the fact that the stability–plasticity balance of the network is highly de-coupled from the all other parameters; and (iii) improved statistical coherence in comparison with Fuzzy ART.

It is believed that (i) SmART has extended the limitation imposed by the stability–plasticity tradeoffs in ART-type architecture; and (ii) the proposed approach of implementing the lateral plasticity control has elegantly solved the major stability problem associated with the use of bi-directional (Euclidean) learning for ART1-type algorithm. Further, SmART learning enables a wide range of plasticity while still retaining good stability. This is made possible by the novel and unique implementation of lateral control of

Table 5  
Suggested values of parameters for SmART with specified vigilance on images of varying complexity

Image colour feature	Plasticity parameter	Learning rate
Simple-‘Objects’ image	Minimal effect	Low
Medium-‘Fruits’ image	High	Low

Table 6  
Comparison of SmART and ART2 in the segmentation of 'Girl' image

Learning rate $\beta$	SmART				ART2			
	No. of clusters	Epochs	Time (s)	MSE	No. of clusters	Epochs	Time (s)	MSE
0.001	20	4	2	419.78	14	3	2	4960.59
0.005	23	6	2	355.81	22	4	3	2963.95
0.01	26	11	3	326.96	31	10	8	2624.17
0.05	42	22	7	280.61	28	8	9	2729.62
0.1	67	57	26	227.92	26	4	4	2768.46
0.5	60	18	8	581.21	–	–	–	–
1	80	28	16	569.23	–	–	–	–

SmART is set at a vigilance of 0.3 and plasticity of 0.8; ART2 is set at a vigilance of 0.995 and noise suppression of 0.4. At the learning rate of 0.5 and above, ART2 failed to meet the stopping criteria.

plasticity in which the adaptivity of each individual prototype is governed by its own internal rules in order to ensure the overall stability of the map. Experimental simulations have verified the theoretical properties of SmART.

The SmART network has also been characterized for use in colour image segmentation applications in which the commitment of the new nodes is indirectly regulated by the plasticity parameter in a manner that is coherent with the complexities of an image. SmART has been found to perform well in *RGB* colour space.

Some possible improvements over SmART to be explored are: (i) generalization of the topological map to arbitrary dimension as suggested by Fritzke [21] for the SOM model; (ii) use of different inhibitory functions (possibly higher-order functions) for the degree of freedom; (iii) use of multiple neighborhood links (instead of just the shortest link); and (iv) use of the angles subtended between, for instance, pairs of links in order to compute their degrees of freedom.

## Appendix A. Appendix A-colour spaces

As is well known, the human visual system (HVS) segments a colour image effortlessly. It can discern thousands of colour shades and intensities, compared to about 16 shades of gray. Owing to the structure of the human eye, all colours are seen as variable combinations of the three so-called primary colours, red (R), green (G) and blue (B). A colour model specifies colours quantitatively by employing a 3D coordinate system and a subspace within that system where each colour is represented by a single point. Some of the more commonly used colour models in image processing are *RGB*, *YIQ*, *HSI*, *Lab* and *Luv*.

In the *RGB* model, each colour appears as its primary spectral components of red, green and blue. This model is based on the Cartesian coordinate system, and is commonly used for representing digital colour images. However, coordinate systems related to the HVS perceptual attributes of luminance, hue, and saturation are often more suitable for processing colour images.

It is desirable that the perceptual difference of colours is proportional to the distance between them in an appropriate colour space. Although the *RGB* model is good for the acquisition or display of colour information, it is not helpful in explaining the human perception of colours. It has been found that (i) if we employ the colour spaces *Lab* and *Luv*, colour differences can be determined by the Euclidean distance measure; and (ii) colour features in the *Luv* space produce more readily separable clusters of pixels.

For the segmentation of colour images, *Luv* space, which has the property of an approximately uniform perceptual space, is preferred because it is associated with a chromaticity diagram in which an additive mixture of two arbitrary colours lies on the straight line joining the two colours. The employment of a perceptually uniform colour space may be advantageous if we try to mimic human performance.

## References

- [1] R. Schalkoff, Pattern recognition: Statistical, Structural and Neural Approaches, Wiley, New York, NY, 1992.
- [2] T. Kohonen, Self-Organizing Maps, second ed. Springer Series in Information Sciences, 1997.
- [3] T. Kohonen, The self-organizing map, IEEE Proceedings 78 (1990) 1464–1477.
- [4] G.A. Carpenter, S. Grossberg, The ART of adaptive pattern recognition by a selforganizing neural network, Computer (1988) 77–88.
- [5] G.A. Carpenter, S. Grossberg, A massively parallel architecture for a self-organizing neural pattern recognition machine, Computer Vision, Graphics, and Image Processing 37 (1987) 54–115.
- [6] A.H. Dekker, Kohonen neural networks for optimal colour quantization, Network: Computation in Neural Systems 5 (1994) 351–367.
- [7] J. Lampinen, E. Oja, Clustering properties of hierarchical self-organizing maps, Journal of Mathematical Imaging and Vision (1992) 261–272.
- [8] H.G.C. Traven, A Neural Network Clustering Algorithm, and its Application to Multispectral Satellite Image Classification Proceedings of the Sixth Scandinavian Conference on Image Analysis, Oulu, Finland, June, 1989 pp. 128–135.
- [9] S. Ghosal, R. Mehrotra, Application of Neural Networks in Segmentation of Range Images Proceedings of the International Joint Conference on Neural Networks, Baltimore, USA, 1992 pp. 297–302.

- [10] N. Papamarkos, E. Atsalakis, C. Strouthopoulos, Adaptive color reduction, *IEEE Transactions on Systems, Man, and Cybernetics* 32 (2002) 44–56.
- [11] T. Uchiyama, M.A. Arbib, Color image segmentation using competitive learning, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16 (1994) 1197–1206.
- [12] S.H. Ong, N.C. Yeo, K.H. Lee, Y.V. Venkatesh, D.M. Cao, Segmentation of color images using a two-stage self-organizing network, *Image and Vision Computing* 20 (2002) 279–289.
- [13] G.A. Carpenter, S. Grossberg, ART2: self-organization of stable category recognition codes for analog input patterns, *Applied Optics* 26 (1987) 4919–4930.
- [14] G.A. Carpenter, S. Grossberg, D.B. Rosen, Fuzzy ART: fast stable learning and categorization of analog patterns by an adaptive resonance system, *Neural Networks* 4 (1991) 759–771.
- [15] H.C. Vlajic, H.C. Card, Vector quantization of images using modified adaptive resonance algorithm for hierarchical clustering, *IEEE Transactions on Neural Networks* 12 (2001) 1147–1162.
- [16] N.C. Yeo, Color Image Segmentation Using Neural Networks, National University of Singapore, M.Eng. Thesis, (1998).
- [17] S. Schunemann, B. Michaelis, Data analysis of not well separable clusters of different feature density with a two-layer classification system comprised of a SOM and an ART2-A network Proceedings of the IEEE International Conference on Neural Networks, Anchorage, USA, 1998 pp. 707–712.
- [18] D.M. Cao, Color Image Segmentation Using Neural Networks, National University of Singapore, M.Eng. Thesis (1996).
- [19] T. Frank, K.F. Kraiss, T. Kuhlen, Comparative analysis of fuzzy ART and ART2A network clustering performance, *IEEE Transactions on Neural Networks* 9 (1998) 544–559.
- [20] B. Fritzke, Let it Grow-Self-organizing Feature Maps with Problem Dependent Cell Culture Proceedings of the International Conference on Artificial Neural Networks, Helsinki, Finland, 1991 pp. 403–408.
- [21] B. Fritzke, Growing Cell Structures—Self-organizing networks in k dimensions Proceedings of the International Artificial Neural Networks, Brighton, UK, 1992 pp. 1051–1056.
- [22] B. Moore, ART1 and pattern clustering Proceedings of the Connectionist Models Summer School, San Mateo, USA, 1988 pp. 174–185.