Intelligent Image Inpainting based on a Brain-Mimicking Recognition-Mining-Synthesis Network

Shao-Han Tang[1], Chuan-Yung Tsai[2], Yu-Han Chen[2], and Liang-Gee Chen[1, 2]

[1]Department of Electrical Engineering and [2]Graduate Institute of Electronics Engineering, National Taiwan Univ., Taipei, Taiwan

Abstract—**In this paper, a new image inpainting method based on a brain-mimicking Bayesian network is proposed. It features the Hierarchical-Spatial-Temporal prediction for simulating brain's neural projections. The trained network can *recognize* image and automatically *mine* for defects by its attention model, before *synthesizing* lost image parts. Difficult inpainting problems, like images with important features lost, are shown to have good results.**

I. INTRODUCTION

Inpainting, the reconstruction of lost image parts, is an important multimedia processing technique. However, just like most of the other intelligent multimedia applications, image inpainting still has much room for improvement, despite being widely researched.

In 2005, Intel proposed a workload model for all next generation applications, called the Recognition-Mining-Synthesis (R-M-S) model [1]. They can be defined as simple questions: *R*—what is it? *M*—where is it? *S*—what will it be? It can be inferred that most intelligent multimedia applications comprise R-M-S problems (e.g. inpainting at least needs *S*), and giving a general solution to R-M-S problem is the most significant yet most difficult challenge. Nevertheless, we know that our brain is the best R-M-S solver. Actually, there does exist the *what pathway*, *where pathway*, and *planning-action pathway* in human brain, which can be mapped to *R*, *M*, and *S* functions. Therefore, we believe deriving a brain-mimicking unified algorithm is best suitable for all the R-M-S problems, since the neocortex consists of solely homogeneous structures too.

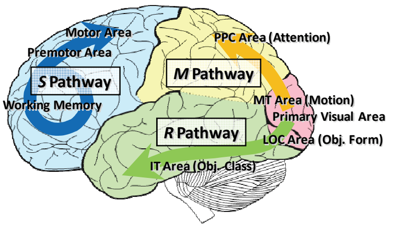


Figure 1. R-M-S pathways in human brain

In this paper, we show that the proposed R-M-S network can do inpainting by synthesizing lost image parts. The object recognition and defect mining functions are also featured in this new method.

II. PREVIOUS WORKS

We classify previous inpainting methods into 2 categories.

A. Intra-image Inpainting

Intra-image inpainting propagates the texture and structure from the known parts to reconstruct the defective parts. Bertalmio [2] first proposed PDE-based method to propagate textures with maximized continuity. Exemplar-based method uses texture comparison and replication to insert texture patches into the defective parts [3]. Sun [4] proposed structure propagation by human assistance (drawing the line of structures in the missing parts) which deals with the hard problem of restoring complex structures.

B. Inter-image Inpainting

Inter-image inpainting reconstructs the defective parts by similar images in its database. The million of photographs [5] method can fill in large defective part which was too large to be propagated in traditional ways. Also, it can fill in the defective part using novel textures which are not shown in the defective image.

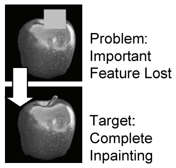


Figure 2. Example: apple

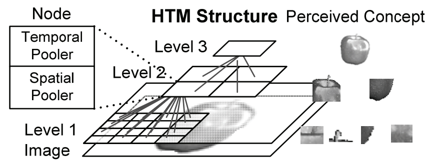


Figure 3. Overview of HTM network

Nevertheless, for a defective image which loses its important feature (like Fig. 2), it's impossible to reconstruct this feature (the apple stem) by intra-image inpainting. Inter-image inpainting may be able to solve this hard problem if its database contains apple images. However efficient image representation and searching for its database are even harder problems. As a result, we try to propose a new algorithm as the total solution for these problems.

III. ALGORITHM

A. Reference Algorithm: Numenta HTM Network

Hierarchical Temporal Memory (HTM) [6] models the neocortex into a hierarchical structure as shown in Fig. 3. As the hierarchy goes up, the more complete object parts can be memorized. Another important concept is the temporal grouping. It means that object parts which appear together will be grouped together. In a node (mimicking the cortical column), spatial pooler decides what pattern the signal is, and temporal pooler decides what group the pattern is. All the beliefs propagated from spatial pooler to temporal pooler and node to node form a Bayesian network. However the propagation direction is upward only in current HTM software.

B. Proposed Algorithm: R-M-S Network

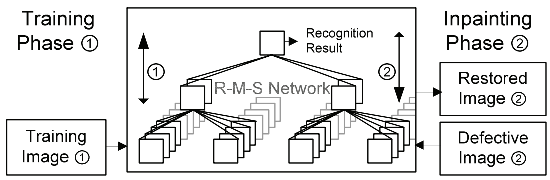


Figure 4. R-M-S network and its working flow

Fig. 4 shows the proposed R-M-S network. First, a supervised training for image recognition is needed. When doing the unsupervised inpainting, the defective image is recognized first, and then the recognition result is propagated back through the R-M-S network to find out the defective parts and predict their original patterns. The major new features are described as follows.

1) Hierarchical-Spatial-Temporal (HST) Prediction

In addition to the upward belief propagation (BP), we propose the HST prediction, as network's new BP directions shown in Fig. 5, to simulate the neural projection directions. *H* prediction sends beliefs from higher (neocortical) levels. *S* prediction sends beliefs from 4 neighbor nodes in the same level. *T* prediction models the time-delayed thalamic feedback of this node's belief. Their major advantages for inpainting are: *H*—deliveing the whole object's concept to lower levels for reasonable texture selection, *S*—preserving object parts' spatial correlation, and *T*—stabilizing the belief's transition. With the HST prediction, a powerful basic structure of the R-M-S network is realized.

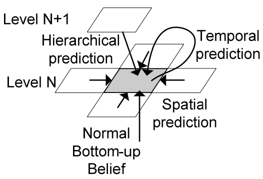


Figure 5. HST prediction

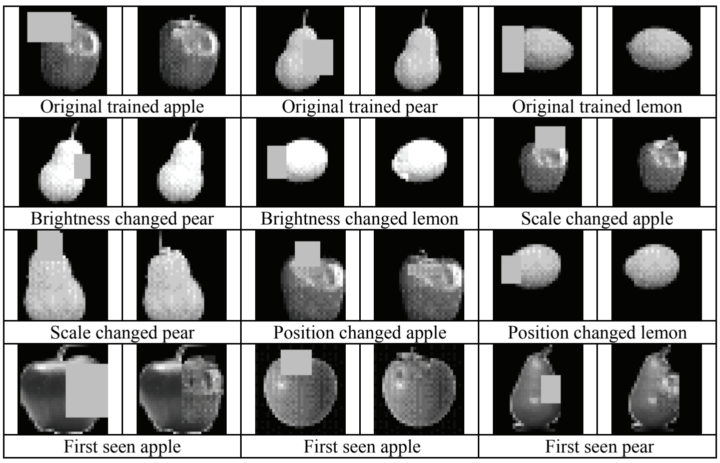
2) Attention Model: Mining for Defects

The visual attention can be bottom-up or top-down. For example, significant color contrast which catches our attention is bottom-up. When seeing defective images, top-down attention pops out due to the strong difference between *what we see* (bottom-up beliefs) and *what we expect to see* (majorly *H* prediction). Using this top-down attention model, the network itself can find out the odd parts and do inpainting by gating the bottom-up beliefs, and let HST prediction to predict and fill in (i.e. synthesize) the lost beliefs (patterns). When this process ends at the network's bottom level (image level) and all BPs converge, the inpainting finishes.

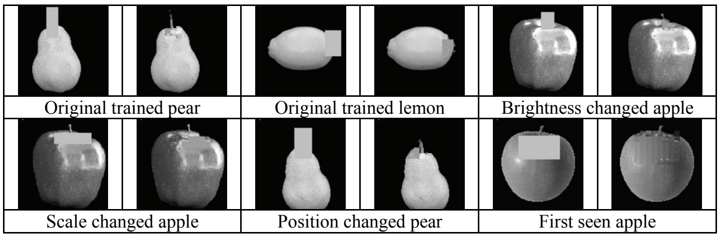
IV. Results and Comparison

The R-M-S network is developed based on HtmLib [7]. Three-category (apple/pear/lemon images) test cases are used to verify its performance, with important features (like stems or near-stalk parts) being purposely masked. Averagely, 32x32 results are good. But the smoothness of 128x128 results are not very satisfying—though important features are restored well, yet the reconstructed textures still need further processing. Hybrid method like R-M-S network guiding a traditional inpainting unit can improve this too.

A. 32x32 Test Case



B. 128x128 Test Case



C. Feature Comparision

Table 1 summarizes the comparison between proposed method and previous works. For novel texture restoration, our method can outperform inter-image inpainting since the images can be recognized before synthesizing reasonable textures. Although our storage size can be very large (around 30MB for 32x32 test case), we believe it will be feasible soon with the rapid growing technologies. Most importantly, attention model enables the possibility of image inpainting without any user intervention.

TABLE I. Qualitative Feature Comparison

Method \ Feature	Novel Texture Restoration	Storage Size	User Intervention
Intra-Image	No	Very Small (inpainting rules)	Partial (mark defects) or Yes (ref. [4])
Inter-Image	Partial (may not be reasonable)	Large (image database)	
Proposed	Yes	Very Large (trained network)	Partial (mark defects) or No (use attention)

V. Potential Application

Intelligent image inpainting based on R-M-S network can have wide range of applications. For example, in the emerging 3D TV technology, the object occlusion problems during 2D-to-3D video source conversion can suitably employ our method. Digital TV's error concealment is a good target application too. Practically, it can provide fully automatic processing for any related problem.

VI. Conclusion

In this paper a new method for image inpainting based on the brain-mimicking R-M-S network is proposed. With the HST prediction, the image is recognized and its defects are mined out by attention model, before synthesizing the lost image parts. Unlike previous works, even images with important features lost can be handled with our method. The brain-mimicking R-M-S network also gives a new point of view to solve inpainting, and other intelligent problems. We believe it can be widely used in the future.

VII. Reference

[1] P. Dubey, "A platform 2015 workload model: Recognition, mining, and synthesis moves computers to the era of tera," Intel Corp. White Paper, 2005.

[2] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," Proc. of the ACM SIGGRAPH, 2000.

[3] A. Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," IEEE Trans. on Image Processing, 2004.

[4] J. Sun, Y. Lu, and J. Y. Jia, "Image Completion with Structure Propagation," Proc. of the ACM SIGGRAPH, 2005.

[5] J. Hays and A. A. Efros, "Scene completion using millions of photographs," Proc. of the ACM SIGGRAPH, 2007.

[6] J. Hawkins and D. George, "Hierarchical Temporal Memory—Concepts, Theory, and Terminology," Numenta Inc. Whitepaper, May, 2006.

[7] G. Kochaniak, *HtmLib 7*, 2008.