A framework of image understanding based on PDP model. Proposition of ICE (image centered) system

Tsuruta, Naoyuki
Department of Intelligent Systems, Kyushu University

Akagi, Takuma
Department of Intelligent Systems, Kyushu University

Taniguchi, Rin-ichiro
Department of Intelligent Systems, Kyushu University

Amamiya, Makoto
Department of Intelligent Systems, Kyushu University

出版情報：Proceedings of 1993 International Joint Conference on Neural Networks, IJCNN '93-Nagoya. 1, pp.177-180, 1993-10. IEEE
バージョン：

Proceedings of 1993 International Joint Conference on Neural Networks

A Framework of Image Understanding Based on PDP Model -Proposition of ICE(Image CEntered) System-

Naoyuki TSURUTA, Takuma AKAGI, Rin-ichiro Taniguchi and Makoto Amamiya
Department of Information Systems
Kyushu University
Address:6-1, Kausga-Koen, Kasuga, Fukuoka 816 JAPAN
E-mail:tsuruta@is.kyushu-u.ac.jp

abstract

One of the most important tasks in image understanding is to map visual information into symbolic concepts which describe an input scene. However, the mapping presents the following difficulties:

1.How to resolve the ambiguity in visual information.
2.How to reduce the redundancy of visual information.
3.How to describe scenes efficiently using symbolic concepts.

In this paper, we propose the ICE System, which is a framework of computer vision addressing the above three problems. First we will see a multi-layered model based on the hypercolumn with selective attention mechanism can solve the first two problems. Then, we will describe how the ICE System is structurally constructed on the basis of the multi-layered model.

1.Introduction

One of the most important tasks in image understanding is to map visual information into symbolic concepts which describe an input scene. However, the mapping presents the following difficulties:

1.*How to resolve the ambiguity in visual information.*
Because understanding the three-dimensional world from a two-dimensional image is an "ill-posed problem", vision problems always generate many ambiguous solutions. In general, to derive an unambiguous output, many natural constraints must be considered and applied.

2.*How to reduce the redundancy of visual information.*
Vision must derive the same symbolic concept for the same object regardless of its position.

3.*How to describe scenes efficiently using symbolic concepts.*
Since the image contains a lot of information, it is difficult to represent all the information simultaneously by symbolic concepts.

For the first difficulty, several studies have employed an approach whereby the result of image understanding is considered to be an optimal solution that satisfies all the multiple simultaneous constraints. The redundancy problem, however, is not fully considered in this approach[1,2,3,4,6,7].

For the second difficulty, the multi-layered model based on the hypercolumn structure is useful, and furthermore, with its mechanism of selective attention, can also solve some ambiguity problems. But it cannot describe the complex relations among symbolic concepts. To solve the third difficulty, a combination of the appropriate distributed representation and the sequential understanding process is necessary[8]. However, there had been no attempt to solve all three problems simultaneously.

In this paper, we propose the ICE System, which is a framework of computer vision addressing the above three problems. First we will see a multi-layered model based on the hypercolumn with selective attention mechanism can solve the first two problems. Then, we will describe how the ICE System is structurally constructed on the basis of the multi-layered model.

2.Multi-layered Network Model

To reduce the redundancy of visual information, multi-layered model based on the hypercolumn structure, is useful:in its bottom up process, a feature extraction process and a feature integration process are iterated alternatively. The feature extraction process extracts complex and global features from activity patterns of simple and local features at a lower level, while the feature integration process extracts abstracted features by combining lower features. The reduction of location redundancy in Neocognitron[5] is one of feature integration process.

Multi-layered model is also useful to resolve the ambiguity in visual information using higher-level (global) information. Fig.1 shows the multi-layered model with selective attention mechanism. Two adjacent hypercolumns in a layer U are denoted by C_1 and C_2. In this case, C_1 and C_2 have receptor fields covering almost the same area (but at slightly different positions) in a lower-layer V. However, if V accepts an input pattern described in the lower part of Fig.1, this pattern is only slightly degraded in C_1's receptive field, but it is degraded much more in C_2's receptive field. So, in the initial state, the "L vertex" unit is almost fully activated in C_1, but no unit is activated in C_2 (See the upper part of Fig.1). Then, after the computation according to the activation rule, the "line" unit in C_2 will be activated because the input pattern is reconstructed in C_1's receptive field, and finally, the input pattern is also reconstructed in C_2's receptive field (See the lower figure in Fig.1).

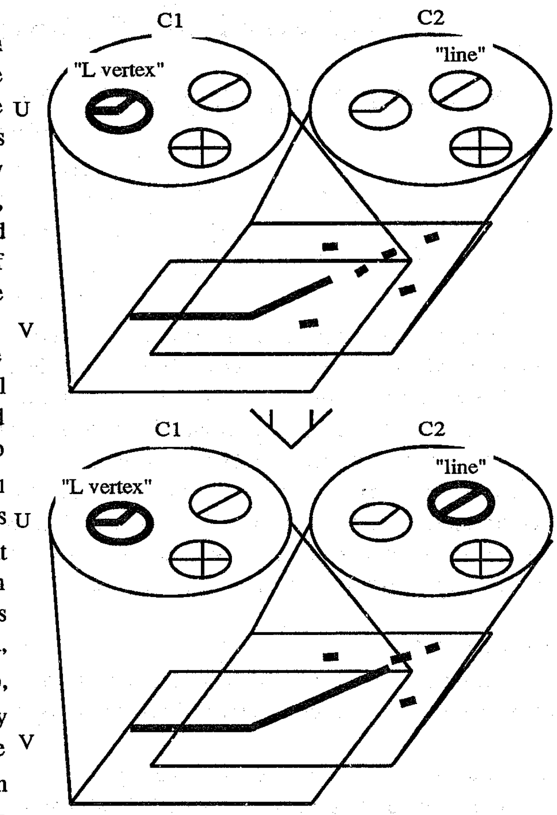

Fig.1 How to reduce the redundancy in multi-layered model.

Thus the multi-layered model with selective attention mechanism is useful to reduce the redundancy and to resolve ambiguity. But, it cannot describe the complex relations among symbolic concepts, because it needs unavailable number of units to represent each relation with one unit respectively. To overcome this problem, in next section, we will propose the ICE System which is structurally constructed on the basis of the multi-layered model.

3.Structure of ICE System

In the ICE System, image understanding is considered to be a process which maps visual information to symbolic concepts, which are each defined by different attributes. The system represents the result of visual processing in a set of concepts, with each concept corresponding to the most activated unit in each attribute. The structure of the ICE System is shown in Fig.2. Now, we propose to use parts of speech as the attributes of a highest layer. This is similar to the representation of an image in a simple sentence.

The lowest layer of the ICE System accepts a reconstructed image from an early vision system. The intermediate layer construction is based on the multi-layered cooperative model. However, if the size of an input image is 1000 $\times$ 1000 and the maximum gray level is m, then the number of possible images is $m^{1000\times1000}$. On the other hand, the number of concepts or entries in a dictionary is no more than 10^4-10^5. Then it is indispensable to compress the amount of information. Therefore to overcome this problem, the feature integration process of the multilayered network is quite useful.

As previously mentioned, a scene is described in a set of words, with each word corresponds to the most activated unit in each attribute group simultaneously. Therefore if there are many objects in a scene, the ICE System cannot derive a description or a sentence which represents all the information in the image simultaneously. We should note that this problem arises with any image understanding system, not only with the ICE System, and a sequential understanding process is required to solve the problem.

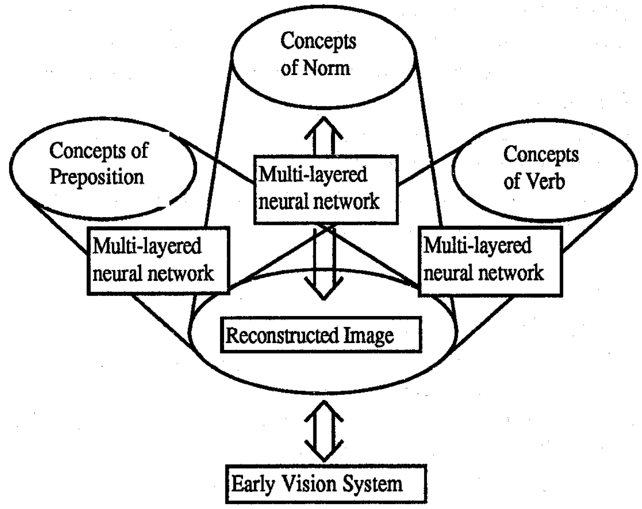

Fig.2 Structure of ICE System

4.Sequential Understanding Process

For the sequential understanding process, selective attention is also the key issue. Once the ICE System focuses on one concept, that is, only one unit is activated in the highest-layer attribute by the competing mechanism, features related to the selected concept are also activated fully down to the lowest layer *(centered image)*, but features having no (negative) relation to the selected concept are inhibited on the lower layers. Therefore, as the result of the cooperative process between the attributes and its intermediate system, only one relation is consistently selected by all the attributes at the same time. The attention transfer can be performed by using a mechanism whereby an activated unit are periodically deactivated and then reactivated. However, since this simple mechanism is not intelligent, we are researching a new higher level system that can control attention strategically.

Within the framework of the ICE System (structure), we can use the cooperative process called Subsymbolic Processing paradigm[8]. Fig.3 shows a simplified example of Subsymbolic Processing between parts of speech, with a distributed representation of the concept of "penguin." In this image analyzing process of ICE System, the concept of "penguin" is activated based on visual features, such as colors, size and shape, of the object and its circumstances, i.e., "an object is on the iceberg." At the same time, a sub-concept "moving" is activated from the image*. Then the sub-concept "moving" activates a concept of"walk," i.e., "moving on the ground," because an object is on the iceberg (or the ground). However, the concept of "fly," i.e., "moving in the sky" cannot be activated, because the concept of "sky" has a negative relation with the iceberg, which is strongly the hypothesized by the image analysis.

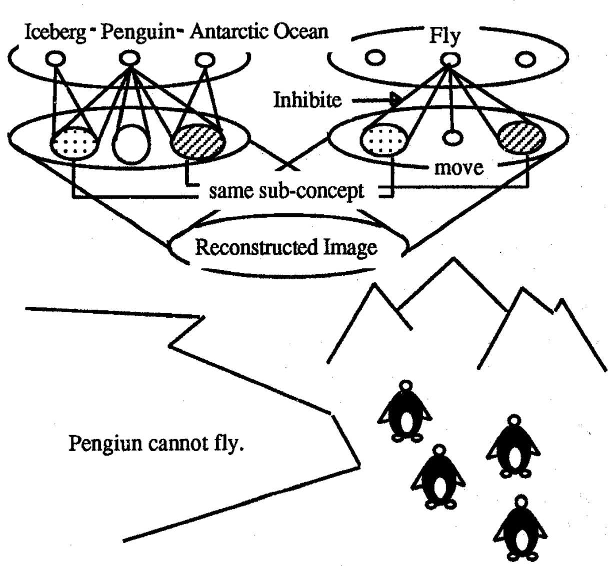

Fig.3 Simplified example for subsymbolic Processing.

* Strictly speaking, it is a sequential of images.

When a knowledge built in ICE System represents that "a penguin is a bird"and "a bird can fly," the activation of "penguin" activates the concept of "bird," and the concept of "bird" activates the concept of "fly." In this case,there occurs a contradiction, i.e., "a penguin is walking (not flying) on the iceberg" vs. "a penguin (a bird) flies." When the hypothesis acquired by the image analysis is strong enough, i.e.,the activation of the concept of "iceberg" is strong, the activation of the concept "fly" is weakened. However, if the hypothesis of the "iceberg" is not so strong, then the activation of the concept of "fly" becomes active, eventually, the concept of penguin is deactivated when some sub-concepts (circumstances) of "fly" is activated using selective attention mechanism.This kind of Subsymbolic Processing is realized by ICE System.

5.Conclusion

Mapping visual information into symbolic concepts presents the following difficulties:

1.How to resolve the ambiguity in visual information.

2.How to reduce the redundancy of visual information.

3.How to describe a scene efficiently using symbolic concepts.

The multi-layered networks, which are based on the hypercolumn can solve the first two problems, but they do not have any effective structure for the third problem, because they have neither a distributed representation mechanism for complex scenes nor a mechanism to maintain the coherency between the attributes in the distributed representation. In contrast with the ordinary multi-layered networks, the ICE System, which is structurally constructed on the basis of the multi-layered cooperative networks, can solve the third problem. The key feature of the ICE System is that it consists of multi-layered networks, each of which corresponds to an extraction process of a part of speech concept. This structure provides cooperative processing among extraction processes of part of speech concepts, and maintains the consistency of the information represented in the entire network quite efficiently. We believe that this feature of the ICE System is indispensable for complex image understanding, in which quite a large amount of information, extracted from visual input and given by the knowledge of the world, should be represented in a large-scale network. From this point of view, the ICE System provides a good framework of massively parallel computer vision.

At present, we are constructing extraction networks for prepositional and verbal concepts in the ICE System, and also we are making further experiments to clarify the effectiveness of Subsymbolic Processing.

references

[1]T. Poggio, V. Torre and C. Koch, "Computational Vision and Regularization Theory," *Nature*, 317, pp.303-323(1985).

[2]M.Kawato,T.Inui, "Computational Theory of the Visual Cortical Areas," *IECE D-II*, Vol.J73-D-II, No.8, pp.1111-1121 (1990)

[3]B. Gidas, "A renormalization group approach to image processing
problems," *IEEE Trans Pattern Anal. & Mach. Intell.*, Vol. PAMI-11, No. 2, February, pp.164-180(1989)

[4]H.Igarashi,M.Kawato, "An Extension of the Standard Regularization Theory," *IEICE Technical Report*, NC91-58(1991)

[5]K.Fukushima, "A neural network model for selective attention in visual pattern recognition and associative recall," *Applied Optics*, 26[23], pp.4985-4992(1987)

[6]S. Gemman and D. Gemman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images", *IEEE Trans Pattern Anal. & Mach. Intell.*, Vol. PAMI-6, No. 6, November, pp.721-741(1984)

[7]K.Kakusho,T.Kitahashi, "Three Dimensional Interpretation of an Image based on Energy Minimization," *IEICE Technical Report* PRU89-108(1990)

[8]D. E. Rumelhart, J. L. McClelland(Eds.), *Parallel Distributed Processing : Exploration in the Microstructure of Cognition*, Vol. 1, 2, MIT Press, Cambridge(1986)