

# 相関情報抽出ネットによる ステレオ画像上物体の奥行き情報抽出の教師なし学習

Unsupervised Learning to Extract the Depth of an Object on a Stereo Vision  
Using the Correlated Information Extracting Neural Network

柴田 克成 (PY)、岡部 洋一 東京大学 先端科学技術研究センター

Katasunari Shibata (PY) and Yoichi Okabe

Research Center for Advanced Science and Technology (RCAST), Univ. of Tokyo

shibata@okabe.rcast.u-tokyo.ac.jp, okabe@okabe.rcast.u-tokyo.ac.jp

**abstract-** Correlated Information Extracting Neural Network has been proposed to extract the common information among multiple kinds of inputs, such as visual and motion signals without any supervisor signals. Using this neural network, the depth of an object on a stereo vision could be extracted after learning.

## 1. はじめに

複数種類のセンサ情報を統合、融合することにより、単に情報の精度を上げるという量的な効果だけでなく、センサ情報処理の質的な飛躍が望まれている。複数種類のセンサおよび自らの運動の情報の中に共通に存在する情報は、我々生体の生活に欠かせない情報である。例えば、我々が前に進めば、目に映っている物体は大きくなり、やがて手に届くようになる。この因果関係、つまり我々が生活する3次元空間の認識を学習することによって、我々は、食物が目の前に見えたら前に進み、手を伸ばして、食べることができる。

また、Heldらの実験[1]によれば、図1のように、2匹のネコが、同じ視覚体験をしても、箱に入れられ、自らの意志で動作できないネコは正常な空間認識能力を形成することができない。これは、空間認識能力の形成には、視覚の情報だけでなく、自らの意志による運動が必要であるためと考えられる。また、逆に、視覚と運動の情報から空間認識能力が形成できる可能性を示している。

複数種類の情報から、その中に共通して存在する情報を抽出することを学習するための方法として、類似した2つの方法が独立に提案されている[2][3]。また、後者の方法(相関情報抽出ニューラルネット)を用いて、動作と視覚の情報から物体までの相対位置を教師なしで学習できることが示されている[4]。しかし、そこでは、視覚センサが一つであるため、物体の大きさ固定という条件を付加していた。

そこで、ここでは、2つの視覚センサを用意し、得られたステレオ画像から、物体の大きさによらない視覚センサから物体までの距離を相関情報抽出ネットを用いた学習によって抽出させることを試みる。

## 2. 相関情報抽出ニューラルネット

相関情報抽出ニューラルネットは、図2のように、2つの階層型ニューラルネットによって構成される。そして、それぞれのニューラルネットに、別々の種類の情報を入力する。例えば、片方に視覚に関連する情報、残りの片方に運動に関連する情報といった形である。そして、片方のニューラルネットの出力をもう片方のニューラルネットの教師信号として、互いに出力を交換しあい、両者の出力が同じ値に近づくようにバックプロパゲーション法に基づいて学習を行う。これによって、両出力は、両方のニューラルネットへの入力に共通した情報を入力するようになる。また、出力の値が、値域の中に広く分布するように、入力の提示何回かに1回(ここでは、100回)、その提示の間で両出力の和が最大の場合と最小の場合に対し、それぞれ0.9、0.1の教師信号を両出力に対して与え、学習を行う。そして、これを繰り返していく。

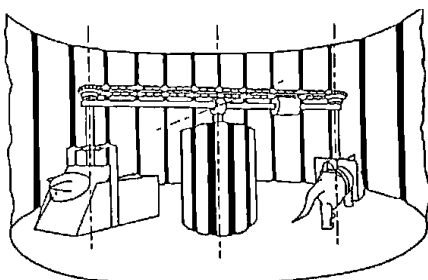
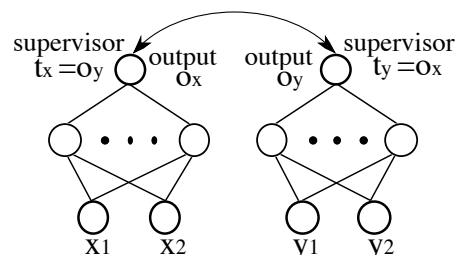


Fig.1 Held and Hein's experiment[1]



Network X (Motion Net) Network Y (Vision Net)

Fig. 2 The structure of the correlated information extracting neural network

[キーワード] 相関情報抽出、センサ情報統合、ステレオ画像、空間認識、ニューラルネット

### 3. シミュレーション

図3のような、2つの視覚センサーを持つロボットを仮定する。このロボットは、2つの駆動部  $\square$ 、 $X$  を持ち、6個ずつの網膜細胞を持つ左右2つの視覚センサによって、計12個の視覚センサ信号が得られる。そして、物体の長さ  $l$ 、幅  $t$ 、水平方向の位置ずれ  $z$  が可変であるとする。そして、 $\square$ 、 $X$ 、 $l$ 、 $t$ 、 $z$  を乱数で決定し、その時のそれぞれの視覚センサに映る像を計算し、各網膜細胞はその受容野中に映る物体の面積の割合を出力するものとする。そして、運動の信号である  $\square$ 、 $X$  を片側のニューラルネット (Motion Net) に、視覚センサからの12個の信号をもう片方のニューラルネット (Vision Net) に入力し、前述のような学習を行う。 $\square$ 、 $X$ 、 $l$ 、 $t$ 、 $z$  を変化させて、学習を繰り返す。

結果を、図4、5に示す。図4は、 $\square$ 、 $X$ 、 $l$ 、 $t$ 、 $z$  を色々に変化させた場合の物体とセンサまでの距離と両ニューラルネットの出力との関係を示した。Vision Net の出力の方が、出力のばらつきが多少あるものの、両ニューラルネットの出力とセンサと物体間の距離がほぼ1対1の対応がとれていると言える。図5は、 $X$  および  $l$  と両ネットワークの出力の平均値との関係を示した。ここでは、 $X$  または  $l$  と出力の1対1の対応関係はない。これらことから、両ニューラルネットの出力は、物体の大きさや、個々のセンサ入力によらない、物体とセンサとの距離を学習を通して抽出することができるようになったと言える。

### 4. おわりに

相関情報抽出ニューラルネットの各ニューラルネットに、ステレオ画像と動作の情報をそれぞれ入力し、相互の出力を相手の教師信号としてやりとりして学習することによって、物体と視覚センサまでの距離の情報をステレオ画像の情報からも、動作の情報からも抽出することができるようになった。

### 参考文献

- [1]Held, R. and Hein, A., "Movement-produced stimulation in the development of visually guided components", J.Comp. Physiol. Psychol., Vol. 56, pp. 872-876, 1963
- [2]Becker, S. and Hinton, G. E., "Spatial coherence as an internal teacher for a neural network", Technical Report CRG-TR-89-7, University of Toronto, 1989
- [3]柴田克成, "バックプロパゲーション法に基づくロボットの学習機能に関する研究", 東京大学大学院工学系研究科機械工学専攻修士論文第6章 (東大機械工学研究報告第24巻M23,69-70), 1989 又は、Shibata, K., "A Neural Network to Get Correlated Information Among Multiple Inputs," Proc. of IJCNN'93 NAGOYA, Vol.3, pp. 2532-2535, 1993
- [4]Shibata, K., "Spatial Recognition Model by Extracting Correlated Information between Vision and Motion Information using Neural-Network", Proc. of IJCNN'93 NAGOYA, Vol.3, pp. 2536-2539, 1993

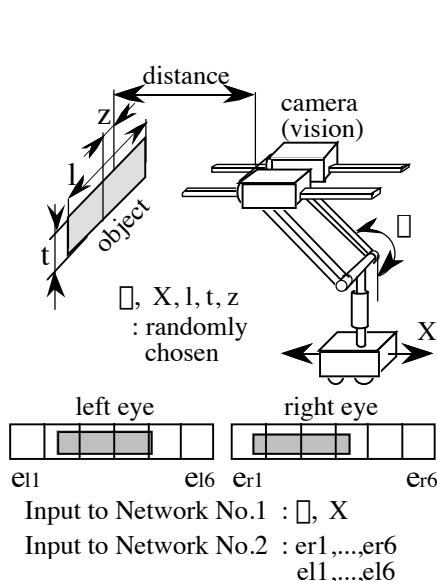


Fig. 3 An assumed robot with two visual sensors

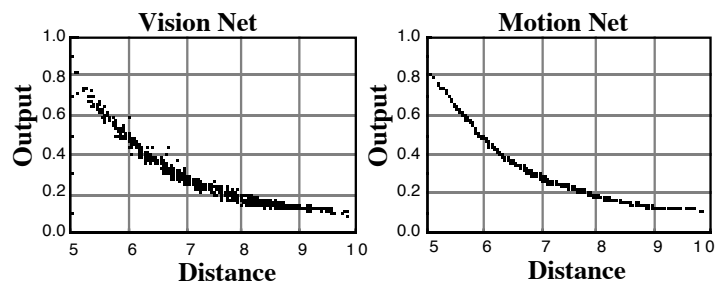


Fig. 4 The relation between the output of each neural network and distance to the object

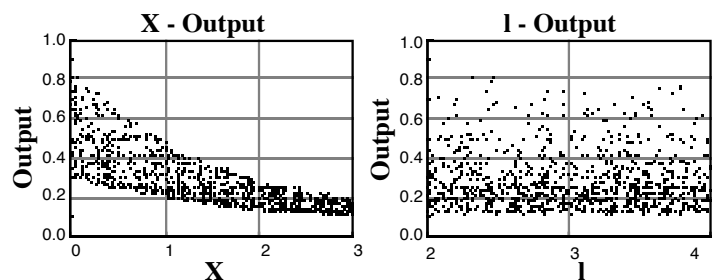


Fig. 5 The relation between the average of two outputs and  $x$  or  $l$