# 應用 GCS 神經網路於 3D 影像重建

計劃編號: NSC88-2213-E-019-001
執行期間: 87 年 8 月 1 日 88 年 7 月 31 日
主持人: 王榮華

## 一、中文摘要 (關鍵字 —GCS 神經網路，動態立體視覺系，競爭式學習，深度圖，向量量化，深度圖)

本計劃研究建立一神經網路動態立體視覺系統(Active Stereo Vision System)。此系統除了能主動的控制攝影機之角度， 使得欲觀察之區域能出現於兩攝影機之影像上，並使用不同類型神經網路來處理影像分割、完成 stereo matching 的相關運算步驟，最後建立此系統可見範圍的精確深度圖。首先，爲改善影像分割之效果，我們使用一個自我發展(self-creating)學習演算法，稱做 ***Growing Cell Structures (GCS)***。每個神經元被賦予一個 resource counter「資源計數器」，以記載其在競爭式學習中贏和輸的機率。此「資源計數器」充分能反映輸入資料之統計特性，使得網路能適時增加或移除神經元。在任何時刻，網路均維持活動力總和爲一個常數，稱之爲資源守恆。由於 GCS 嘗試最大化其熵值，我們稱 GCS 是一個近最佳解的向量量化器。GCS 的自我產生以及可調學習率之機制使得 GCS 能兼顧網路的學習能力與穩定性。在影像編碼的實驗中，我們比較了 GCS 和三種有名演算法 — Kohonen Self-Organizing Feature Map (SOFM) [14]、FSCL [15] 和 SCONN2[16]的效能；模擬結果顯示出 GCS 比其他方法之優越處。

我們進一步將 GCS 運用在一個以神經網路爲基礎的立體視覺系統中。藉由 Sobel 之運算得到立體影像每一點之特徵，如灰階值、變化量和方向性。再將這些特徵給 GCS 進行叢集影像；相同的特徵也用來訓練倒傳遞網路成爲一個匹配器。訓練後的網路匹配器可用來產生一個初始的深度圖。之後，再使用一個以影像叢集爲基礎的對應點演算法，以得到一個更精確的深度圖。 實驗結果顯示出倒傳遞網路匹配器和對應點演算法的成效。

## 英文摘要 (**Key words** — GSC neural networks, Active stereo vision system, Competitive learning, Vector quantization, Disparity map)

A self-creating neural network effective in learning vector quantization, called GCS (*Growing Cell Structures*) is introduced. Each neuron in GCS is characterized by a measure of *resource*. Conservation is achieved by bounding the summed resource of all neurons at a constant, despite value for which varies from one network to another. Resource values of all neurons are updated after each input presentation. We show that GCS effectively fulfills the conscience principle and achieves biologically plausible self-creating capability. In addition, conservation in resource facilitates systematic derivations of learning parameters, including the adaptive learning rate control useful in accelerating the convergence as well as in improving node-utilization. The performance of GCS is compared with three famous algorithms — Kohonen's Self-Organizing Feature Map (SOFM) [14], FSCL [15], SCONN2[16]..

The GCS is further used in a stereo vision system based on neural networks [4]. Sobel operators are used to extract features of intensity, variation, and orientation from stereo image pairs. These features are used to clustering images by GCS and to train a BP neural network in order to obtain an adaptive matcher. The trained BP matcher can generate an initial or primitive disparity map that provides necessary correlation or corresponding SSD (sum of squared differences) in area-based matching methods. Following the BP training, we propose a matching algorithm based on image clustering that is useful in iteratively updating the primitive disparity map. We show that this update algorithm can improve the quality of the disparity map significantly. Empirical results show that the efficiency of the BP matcher and the validity of our matching algorithm.

## 二、計畫緣由與目的

在電腦視覺的領域中， 景深資訊(depth information)的取得一直是個很重要的問題。不論在機器人控制[1]，無人車控制[2]，遙測[3]，醫學影像分析[4]等應用上，系統是否能自主以及精準的測知目標之深度是成敗關鍵。而使用立體視覺[5]系統(stereo vision system)來取得景深資訊一直是個熱門研究領域。其主要概念是使用兩具或以上的攝影機來取得平面影像，再利用已知的資料(如：影像、攝影機的焦距，位置，角度)可直接計算出被觀測點的實際座標及測量目標之景深。

以往，立體視覺的研究 [5,6,7]大多在於如何能從不同的影像中找出相對應的點(registration 或 stereo matching)，一般使用兩具平行架設的攝影機(non-convergent system)來取得平面影像，這固然可以使問題大大的簡化；亦即，一旦得到兩影像間正確的對應，就可由對應點的座標、攝影機的焦距(focal length)、以及攝影機架設的距離這些已知的參數值，透過相似三角形的運算而得到被觀測點的座標。然而此種系統的問題在於攝影機間的距離限制了計算座標之精密度。距離越大可有較高的精密度。但是若加大攝影機的距離，則兩張攝得影像中之對應部份會減少，以致縮小了可計算的區域，其最嚴重者甚至無法觀測目標物。因此，有必要研究主動式視覺系統(active vision system)[8,9]，此類的系統能動態控制攝影機的位置以及方向，使每具攝影機都能拍攝所欲觀測的目標，而使得所攝入之影像容易分析。如此，不僅加大觀測區域，且因提供更多的 cues (如:攝影機 vergence 之角度、攝影機之焦距以及座標等) 而提高座標計算精密度。此外，輸入影像正確分割一直是影像分析的重要課題。影像的分割可使得攝影機能快速的對準同一個區域。精確的分割也可成為建立深度圖(range map)的重要資訊。同時，吾人亦考慮神經網路毋需針對輸入資料的結構關係預作假設，即能學習 unstructured knowledge 、和解決非線性問題之特性，極具實現 real-time 系統的潛力。近來，已有應用神經網路於立體視覺的研究[7]。這些研究集中在嘗試以神經網路來解決對應點匹配的問題。Stereo matching 強調處理立體視覺時找出兩張攝自不同角度影像之相對應的點，此類問題通常可轉換為為最佳化的問題。

本計畫主要是研究利用神經網路去解決下列關鍵性問題：

**影像分割(segmentation)**: 傳統影像分割的技術已成熟者有: (1) Region-based 方法[10]，此類方法概念簡單但需較久的計算時間；(2)Feature-based 方法[11]，則往往需對特徵空間作困難之非線性分割。近來有許多以神經網路應用於影像分割的研究[12]。監督式學習的網路，如倒傳遞神經網路(Backpropagation Neural Network、BPNN) [13]，需要一些事前的資訊來學習，網路學習的速度較慢。非監督式學習的網路，如 SOFM 網路[14]能快速的學習影像特徵之聚類關係，且不需要事前的分割資訊。為改善影像分割效果，我們亦改良完成一個自我組織競爭學習演算法 GCS [17]。每個神經元有一個「資源計數器」、以估測其在競爭學習中贏和輸的機率。此資源計數器能充分反映輸入資料之統計特性，使得網路能適時增加或移除神經元。在任何時刻，網路均維持活動力總和為一常數，稱之為資源守恆原則。這個守恆原則最大優點是幫助我們推導出重要學習參數的設定。增加或移除神經元會影響輸入向量空間之分割，也改變了網路的熵值。在影像編碼的實驗中，我們比較了 GCS 和三種有名演算法 — Kohonen SOFM [14], FSCL [15] 和 SCONN2[16]的效能：模擬結果顯示出 GCS 比其他方法之優越處。

**對應點匹配(stereo matching)**: 這可說是立體視覺中最重要的問題。已發展的對應點匹配技術大多可分為以區域為基礎(area-based)[6]和以特徵為基礎(feature-based)[7]的匹配。以區域為基礎的技術大多使用 Auto-correlation 或是 Sum of Square Difference (SSD)來計算兩張立體影像中相匹配的點。所選取視窗之大小影響匹配結果的優劣。而且 SSD 等方法對於視窗內影像之分佈也無法考慮進去。而以特徵為基礎的技術則是先對影像中特殊的點(points of interest)進行匹配，再使用內差法(interpolation)得到完整的深度圖。但是內差法所得之結果通常無法正確的與實際深度吻合。本系統將結合區域為基礎以及特徵為基礎的匹配方法，以求得更精確的結果。本系統擬使用倒傳遞神經網路來計算兩個視窗之匹配程度，此方法應可改善傳統 SSD 等方法之缺點。本系統中並提出一匹配演算法，依影像中之特徵(像點變化量之強度)依序進行匹配的過程，以減低內差法之誤差。

## 三、研究方法及成果

### A. 影像分割技術

本系統改良Kohonen的SOFM基本網路來進行影像分割。SOFM是屬於非監督式學習，在網路學習的過程不需知道一些事前的資訊。這很適合用於自動化的系統。SOFM網路能有效的學習輸入向量的聚類關係(clustering)。所以對於輸入特徵向量的選擇就相當的重要。先將影像分爲大小等的小方塊作爲分割的單位。每個方塊都產生其特徵向量。本系統初步使用方塊的位置(position)，平均強度(mean magnitude)，平均變化量(mean variation)，以及平均方向(mean orientation)作爲其特徵向量。其中變化量與方向之定義如下：

$f(x, y)$是影像座標(x，y)的強度， 其梯度值定義爲：

$$\nabla f = \begin{bmatrix} \dfrac{\partial f}{\partial x} \\ \dfrac{\partial f}{\partial y} \end{bmatrix} = \begin{bmatrix} G_x \\ G_y \end{bmatrix} \quad (2)$$

梯度值的強度就定義爲$f(x, y)$的變化量。

$$|\nabla f| = \text{mag}(\nabla f) = \left[G_x^2 + G_y^2\right]^{1/2} \quad (3)$$

在點(x，y)的方向$\alpha(x, y)$ 則定義爲：

$$\alpha(x, y) = \tan^{-1}\left(\frac{G_x}{G_y}\right) \quad (4)$$

Sobel Operator 的運算則提供了很好的 $G_x$ 與 $G_y$ 之近似值。基本的 SOFM 網路對於輸入向量的每個值都給以相等之權重，這會造成一些問題；例如：方塊之位置是很重要的資訊，因爲每個物體大都出現在連續區域，但是若太強調此特徵，則塊大而連續之區域會因位置之不同而分割成爲不同之區塊。因此吾人將特徵向量的每個值都乘上一個權重，使得特徵向量的値能依其重要性而有所調整。隨機取出幾組訓練向量來訓練 SOFM 網路，訓練完成之後再將所有小方塊之特徵向量送入 SOFM，即可於 SOFM 之輸出層得到聚類的關係。

### B. 對應點匹配技術

傳統 area-based 的方法如：Auto-correlation 和 SSD 等方法需使用由左右影像選取視窗(window)來計算匹配程度，其效果受限於視窗的大小。若是選取的視窗太大，則視窗內的影像會因爲攝影角度之不同而產生形變，造成錯誤的匹配。若是所選取的視窗太小，又會造成訊號雜訊比(SNR)過高而產生誤判。另外 SSD 等方法皆未考慮到視窗內影像分佈的情形，單單就視窗的值做運算。這也會造成誤判的影響。本系統使用倒傳遞類神經網路(BPNN)來測量兩視窗之匹配程度。將左右影像所取視窗之灰階差，變化量差以及方向差組成網路的輸入向量，訓練一個三層的倒傳遞網路，此網路的輸出則是兩個視窗的匹配程度指標。我們訓練出來的網路具有較 SSD 等方法更好的匹配能力。視窗大小所造成的問題在此應可減輕，這是因爲網路自己會對不同的位置根據實際訓練資料而加以不同的權重；另外網路也可學得視窗內影像分佈的情形；這使得匹配的效果更合理。 一個左影像的點(pixel)，經由上述 BPNN 的計算可在右影像中得到許多高匹配値的對應點，如何由這些點中選取最合理的一組出來可由一些限制來判斷。常用的限制如唯一性(uniqueness)限制，Epipolar Line 限制，順序(ordering)限制，幾何限制，以及鄰近點(local-support)的限制。

### C. 應用 GCS 神經網路影像分割技術

The first interesting property in GCS we study is the conservation of summed $\tau_i(t)$, which refers to the state when the sum of all resource counters in the network remain unchanged. Assume at time $t$ the $i_{th}$ node wins, the sum of all resource counters at time $t$ can be calculated as the follow

$$\sum_{j=1}^{M(t)} \tau_j(t) = (1-\alpha)\left[\tau_i(t) + s\right]$$

$$(1-\alpha)\sum_{i\neq i} \tau_j(t) \quad (5)$$

If the sum reaches saturation, one must have $\sum_{j=1}^{M(t)} \tau_j(t) = \sum_{j=1}^{M(t)} \tau_j(t+1) = Y = constant.$ It follows that

$$(1-\alpha)\left[\sum_{j=1}^{M(t)} \tau_j(t) + s\right] =$$

$\sum_{j=1}^{M(t)} \tau_j(t) = Y, \quad (1-\alpha)\left[Y + s\right] = Y.$ Hence,

$$\sum_{j=1}^{M(t)} \tau_j(t) = \frac{s(1-\alpha)}{\alpha}. \quad (6)$$

That is, GCS in essence employs the well-known leaky integrator law specified in (5) to achieve the conservation property. To verify Eq.(6), we used a 2-D Gaussian of ($\mu$=150, $\sigma$=30) as test input, and the results for $\alpha$=0.05, 0.1, and 0.15 are shown in Fig.1. As can be seen, $\sum_{j=1}^{M(t)} \tau_j(t)$

always quickly saturates in less than 100 ($=\lambda$) input presentations. The saturation value, interestingly, is irrelevant to $\lambda$ and the forms of input distributions.

Fig.2 shows our proposed architecture for implementing an active stereo vision system. The target-selection step involves GCS training and segmenting the input image pair. The outcome of segmentation greatly effects the stereo matching result and hence the disparity calculation. The image pairs are first divided into many small equal-sized squares. Specifically, the left image is used for training GCS, after training the GCS is then used to segment both images. We use five input features i.e., mean magnitude, mean variance, mean orientation, x position, and y position. Note that the features of mean variance, mean orientation are obtained from Sobel operator are also used in the latter matching process.

Traditionally, competitive learning algorithm treat input features equally, in practice, this simple arrangement might cause incorrect clustering results in dealing with various situations. But if we emphasize the position feature too much without any weighting, a large continuous object tends to be incorrectly segmented into several clusters due to the varying of the position. Scaling down the position feature will lower (compared to other features) its effects during the GCS segmentation processing. Therefore, instead of using more powerful feature extractors which requires large amount of computations, we propose a weighted version of input vector. A 256x256 indoor image shown in Fig. 3(a) is applied to test the segmentation performance of GCS. The image is divided into 4096 equal-size squares, each with 4 pixels height and 4 pixels width. One hundred random features vectors are selected to train a GCS that has 4 nodes in the input layer and 4 nodes in the output layer. After training, all of 4096 feature vectors are input to the GCS. Fig. 3(b) illustrates the results after multiplying the feature by [1.0, 1.0, 0.25, 0.125, 0.125]. For comparison, Fig.3(c) shows the resulting segmentation without weighting.

Note that the proposed scaling weight vector in fact enhances the mean magnitude, and comparatively inhibits the orientation features. This arrangement is reasonable if one considers the situation of processing a large smooth area, too much emphasis on orientation can only introduce unwanted noise. The weight vector [1.0, 1.0, 0.25, 0.125, 0.125] is also applied to other images and yield the same good segmentation performance. Fig. 4 verifies this image-independent property. We conjecture there might exist other weight vectors that would work too. Our results, however, indicate that the weight vector [1.0, 1.0, 0.25, 0.125, 0.125] generally reflects the context structure of input images.

**D. 成果發表**

- J. H. Wang and Chung-Yun Peng, "Competitive neural network scheme for learning vector quantization", *IEE Electronics Letters*, Vol.35, no.9, 725-726, 1999.
- J. H. Wang and W. D. Sun, "On the characteristics of growing cell structures (GCS) neural network", To be appeared in Oct. 1999, Vol. 10, No.2, *Neural Processing Letters*.
- J. H. Wang and Chung-Yun Peng, "A novel self-creating neural network for learning vector quantization", To be appeared in Apr. 2000, Vol.11, No. 2, *Neural Processing Letters*.
- J. H. Wang and J. H. Tsai, "Surface reconstruction using self-creating neural networks." EEE International Conf. on on Systems, Man, and Cybernetics, Accepted, Tokyo,Japan, 1999.

## 四、結果與討論

A self-creatingneural network call GCS capable of performing image segmentation has been studied and successfully applied to 3-D image reconstruction. Characteristics of GCS have been explored. The result is believed to be useful in industrial inspection, robot vision, and remote sensing etc. We note that in using the weighted feature vector, a problem may not be ignored in processing a texture-rich scene in which objects can be wrongly segmented into different regions due to the strong texture property. This problem can be solved by using additional feature (e.g. Gabor filtering) in combination with features presented in this report.

## 五、參考文獻


[1] S. B. Marapane and M. M. Trivedi, "Region-based stereo analysis for robitic applications," *IEEE Trans. Syst., Man, Cybern.*, vol. 19, no 6, pp. 1447-1464, Nov.-Dec. 1989.

[2] S. Tsugawa, "Vision-Based Vehicles in Japan: Machine Vision systems and Driving Control Systems," *IEEE Trans. on Industrial Electronics*, vol. 41, no. 4, pp. 398-405, Aug. 1995.

[3] A. Blake, D. McCowen, H. R. Lo and P. J. Lindsey, "Trinocular Active Range-sensing," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, no. 5, pp477-483, May 1993.

[4] L. Jintae and T. L. Kunii, "Model-based Analysis of Hand Posture," *IEEE Comp. Graph. and App.*, vol. 15, no. 5, pp. 77-86, Sep. 1995.

[5] Umesh R. Dhond and J. K. Aggarwal, "Structure from stereo--A Review," *IEEE Trans. Syst., Man, Cybern.*, vol. 3, no. 1, pp. 1489-1510, Nov. 1989.

[6] Takeo Kanade and Masatoshi Okutomi, "A stereo Matching Algorithm with an Adaptive Window: Theory and Experiment," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, no. 9, pp. 920-932, Sep. 1994.

[7] Nasser M. Nasrabadi and Chang Y. Choo, "Hopfield Network for Stereo Vision Correspondence," *IEEE Trans. Neural Networks*, vol. 3, no. 1, pp. 5-13, Jan. 1992.

[8] E. Grosso and M. Tistarelli, "Active/Dynamic Stereo Vision," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, no. 9, pp. 868-879, Sep. 1995.

[9] N. Ahuja and A. L. Abbott, "Active Stereo: Integrating Disparity, Vergence, Focus, Aperture, and Calibration for Surface Estimation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, no. 10, pp. 1007-1029, Oct. 1993.

[10] Y. L. Chang and X. Li, "Adaptive Image Region-growing," IEEE Trans. Img. Proc., vol. 3, no. 6, p. 868-872, Nov. 1994.

[11] Y. Hu and T. J. Dennis, "Texture Image Segmentation by Context Enhanced Clustering," *IEE Proc. Vision, Image and Signal Processing*, vol. 141, no. 6, pp. 413-421, Dec. 1994

[12] H. Kong and L. Guan, "A Self-Organizing Neural Network for Image Segmentation," Proceedings of the 1994 Second Australian and New Zealand Conference on Intelligent Information, pp. 27-31, 1994.

[13] P. Werbos. *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. PhD thesis, Harvard, Cambridge, MA, Aug. 1974.

[14] T. Kohonen. *Self-Organization and Associative Memory*, volume 8 of Springer Series in Information Sciences. Springer-Verlag, New York, 1984.

[15] S. C. Ahalt, A. K. Krishnamurthy, P. Chen, and D. E. Melton, "Competitive learning algorithms for vector quantization," *Neural Networks*., vol. 3, pp. 277-290, Oct. 1990.

[16] I. Choi and S. H. Park, "Self-creating and organizing neural networks," *IEEE Trans. Neural Networks*, vol. 5, no. 4, pp. 561-575, Jul. 1994.

[17] B. Fritzke, "Growing cell structure: a self-organizing network for unsupervised and supervised learning", *Neural Networks*, vol.7, no.9, pp.1441-1460,1994.

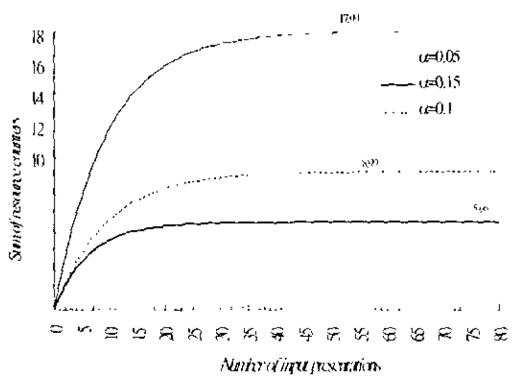



**Figure 1.** Saturation in the total resource counters. $s$=1, $\varepsilon_b$=0.06, $\lambda$=100.

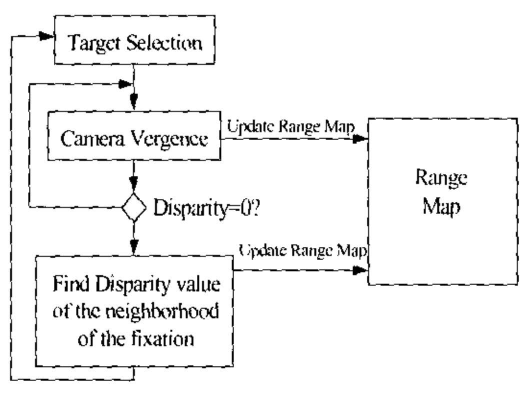
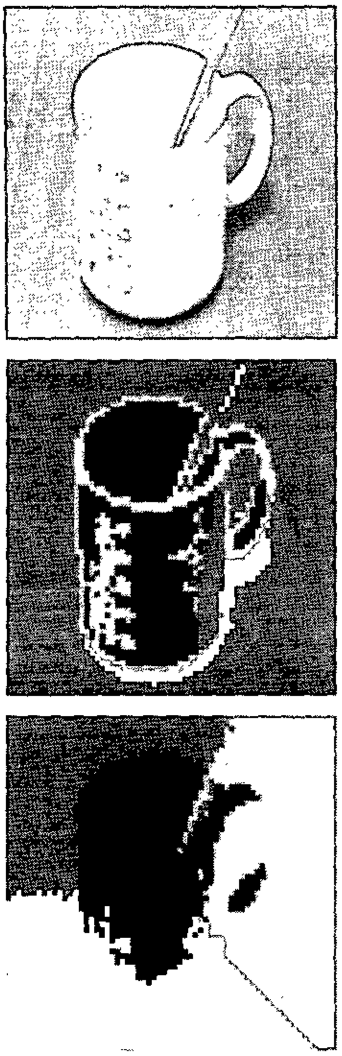


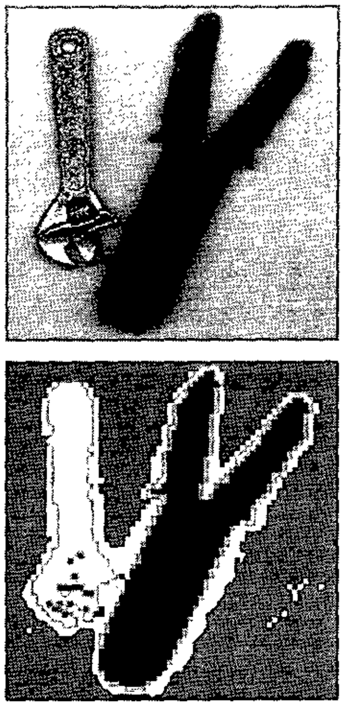

**Figure 2.** The active stereo vision system control flow diagram.



(a)

(b)

**Figure 3**. (a) An indoor image (b) Weighted Segmentation (c) Segmented without weighting.



(a)

(b)

**Figure 4**.(a) Another indoor image -- tools. (b) Weighted segmentation.