

# 精簡報告

## 利用深度學習技術應用於泌乳牛乳頭自動化定位系統之開發

陳建興<sup>1</sup> 吳彥呈<sup>2</sup> 周上詠<sup>3</sup> 張立德<sup>4</sup>

<sup>1</sup> 美和科技大學資訊科技系 助理教授

<sup>2,3</sup> 美和科技大學資訊科技系 學生

<sup>4</sup> 雲林科技大學資訊工程研究所 碩士生

<sup>1</sup> \* jameschen@meiho.edu.tw <sup>4</sup> cowboss779@gmail.com

### 摘要

本研究以深度學習 Yolo V3 演算法的影像辨識技術進行目標物件的定位，透過立體視覺技術量測物件距離，再結合以 Borland C++ Builder 程式開發並完成使用者操作介面以及六軸機械手臂的控制程式，使物件影像辨識、立體視覺測距以及機械手臂動作控制，完全在 Borland C++ Builder 環境下完成，其整合了機械手臂動作控制、深度學習物件辨識以及立體視覺技術，實驗內容是測試系統可對放置在不同位置的乳牛乳頭模型物件進行自動定位與完成機械手臂可自動將擠乳杯套入乳頭模型的動作，目前研究結果發現影像深度在 20cm 至 35cm 的範圍內，機械手臂皆可成功將擠乳杯自動套上乳頭模型，若超出此範圍，則成功率會降低，主要原因是以立體視覺技術所偵測出的影像深度的誤差會隨著實際距離增加而變大，使得擠乳杯套入乳頭模型的成功率受到影響，後續研究將改進影像深度的準確度，使系統能更具有實用性。

**關鍵字:** 立體視覺、深度學習、六軸機械手臂

### Abstract

This research uses the object detection technology with Yolo V3 algorithm to recognize and locate the target object, and measures the image depth of the object with the stereo vision technology, and combines them with the Borland C++ Builder program to develop the user interface and the control program of the six-axis robotic arm. The functions of object detection, stereo vision ranging, and robotic arm control are completely integrated in the Borland C++ Builder environment. The current research results have found that the robot arm can successfully attach milking cup on nipple model within image depths from 20cm to 35cm. If it exceeds this range, the attaching performance will be decreased. The main reason is that the error of the image depth detected by the stereo vision technology will increase through the increase of the actual object distance. The research will further improve the accuracy of the image depth to make the system more reliable and practical in the future.

**Keywords:** 6-axis Robot Arm, Stereo Vision, Deep Learning

---

## 1. 前言

機器人領域正處於一個蓬勃發展的階段。由於演算法的演進和電腦計算能力的提升，使機器人應用由實驗室往其他特殊用途發展，如製造，及進入我們的日常生活[1-3]。這可以從最近的無人自動駕駛汽車技術的興起看出，但機器人技術可能會產生更大的影響。隨著機器人開始從實驗室轉移到現實世界，機器人面臨著許多問題新挑戰。例如考慮一個家庭個人助理機器人，這樣的機器人必須執行許多複雜的任務，例如分揀和折疊衣服、操作電器、拾取和清潔，以及在廚房裡準備食物。此外，它必須處理與之相關的各式各樣的物體及材料等任務，例如，抓取不同的對象，其中一些物件可能是以前未曾見過，或準備不同的食材。對於其中許多問題，在可見的輸入與任務之間只存在一種抽象關係，例如利用視覺資料來控制，或來自觸覺反饋以確定材料屬性。傳統上，人們希望利用手動設計控制器讓機器人來執行的每項任務。即使是人類可直覺完成的任務，例如抓取物體或切割食物對控制器而言是很難直覺地去設計的，因為我們無法輕易地轉譯這種自然直覺為程式代碼。機器人必須在現實世界中處理大量的變化，而擴展這些方法也極具挑戰性，例如 抓取家中的每一件物品，或切割任何食物等。

為了這些理由，近年機械學習演算法已被廣泛應用於機器人[4]，機器學習不是強迫工程師徒手編碼整個端點到端點的機器人系統，而是允許從一些訓練數據中學習系統的每個部分。這種方法允許我們對可能難以或不能正確手動建模的概念進行建模。它還允許適應性模型，一般形式的模型都可以簡單地透過為這些新案例提供訓練數據來適應更多或不同的情況。雖然機器學習算法對於機器人應用具有許多優點，但它們仍然難以應用於新問題。首先，許多學習算法需要耗時的優化來執行推理，使得它們對於具有嚴格時間約束的機器人應用是不可行的。設計模型一般足以應用於同一問題的其他情況，同時仍然具體到足以適應問題，可能非常具有挑戰性，特別是在現實世界機器人中看到的各式各樣的物件。此外大多數機器學習演算法在設計特徵方面需要大量的手動工程-原始輸入的轉換，例如，給予機器人的圖像，及動態訊息等，成為對學習演算法更有用的形式。雖然設計好的特徵對於特定問題的機器學習算法的成功是至關重要的，但是這些特徵通常是針對單一且特定的問題，且需要花費大量的設計成本。

最近，深度學習方法在跨領域表現上廣泛的令人印象深刻[5-10]，包括計算機視覺，音頻處理，自然語言處理等。這些算法基於神經網絡，高度參數化的模型，使用多層表示將數據轉換為特定於任

務的表示。通過使用無監督功能學習演算法，深度學習方法能夠預先初始化這些網絡具有實用功能，避免了常見沒有這種初始化就訓練神經網絡而產生的過度擬合問題。

機器人學為學習算法提出了許多獨特的挑戰。首先是機器人必須執行各種各樣的任務，這通常是耗時的，甚至是無法為每項任務編寫全新的學習演算法和功能。第二，機器人必須處理現實世界中的大量變化，這很困難許多學習算法要處理。最後，時間對大多數機器人應用程序而言都是非常寶貴的，因此學習算法必須適合快速推理對機器人應用程序有用。雖然深度學習是一類廣泛的學習算法，但有很多不同相關的推理模型，大多數方法學習一組連接權重被神經網絡模型用於推理[11]。一個神經網絡，左側通常由多層人工“神經元”組成。每個神經元都與前一層中的每個神經元有加權連接，或者如果神經元位於最低層，則為網絡輸入。然後是神經元通過將其輸入的加權和通過一些非線性來形成輸出激活函數，此輸出然後被發送到下一層。理想情況下，每個功能層都代表更好抽象輸入數據，使最後一層功能更好某些分類器的表示形式比輸入網絡的原始特徵。雖然上述網絡執行二進制分類任務，但它是很容易換掉頂層分類功能用於其他任務，例如運用用於多級分類的 softmax 函數，或者對於 toplayer 的線性加權回歸功能。這種模組化是神經網絡的主要優勢方法。本研究將使用深度學習技術來辨識影像內的目標物件，利用卷積神經網絡(Convolutional Neural Network)[12]方法針對立體視覺影像的物件特徵進行學習，系統可從影像中辨認出動態目標物件位置，再透過立體視覺技術計算出目標物件的視差，以獲得相對於機械手臂距離的影像深度，使機械手臂能往目標物件移動。

## 2. 研究方法

以影像來量測距離的技術可以區分成單眼視覺(mono vision)技術與立體視覺(stereo vision)技術，單眼視覺技術是使用單一個攝影機來拍攝影像，立體視覺方法是使用兩個攝影機來找出影像的深度圖(depth map)，同時拍攝兩張影像，然後使用影像處理技術來找出視差圖(disparity map)，雖然立體視覺技術相較於單眼視覺技術較為複雜與耗時，但準確度較高，所以本研究將採用深度學習影像辨識與立體視覺技術來作為影像景深的量測，底下將詳細說明相關原理以及機械手臂機構設計與運動分析：

### 2.1 AI 影像辨識方法(AI image recognition method)

過往影像辨識方法在電腦視覺架構下，因無法符合實際場域、背景嚴苛考驗下的低準確率，大部

分辨識產品，例：人流計算、警戒線、電子圍籬、人臉辨識等，皆淪為買硬體送軟體或充當玩具的困境，直至人工智慧，深度學習的問世，影像辨識以卷積神經網路作為演算法架構，克服不同場域環境，大幅提升準確性及實用性。卷積神經網路 (Convolutional Neural Networks, CNN) 是深度神經網路領域的發展主力，其影像辨識流程如圖 1 所示，在影像物件的辨識上具有高度精準的程度，目前主要用於物件目標偵測的深度學習模型有 R-CNN、Fast R-CNN、Faster R-CNN、Mask R-CNN 與 YOLO，也都是以 CNN 的架構為基礎去做延伸。本研究採用 YOLO V3 作為訓練時使用的模型 [13]，Yolo 系列 (You only look once, Yolo) 是物件偵測的類神經網路演算法，以 Darknet 為實作架構 [21]，而 Darknet 是一個用 C 和 CUDA 編寫的開源神經網路框架，此實作該架構具備輕量、依賴少、演算法高效率，在工業應用領域很有價值，例如行人偵測、工業影像偵測等等。

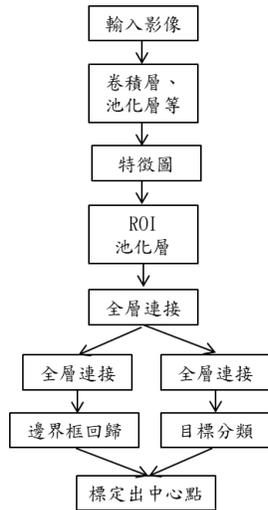


圖 1: CNN 影像辨識流程

## 2.2 立體影像量測方法(stereo image measurement method)

立體視覺技術常被使用於記錄與表示立體空間影像，需要一對攝影機來操作立體視覺，利用在位置略有不同的兩相機各拍攝一張影像就可產生影像深度的視錯覺，兩相機相距固定距離且平行對齊，量測物件距離可以透過兩相機的重疊視圖來完成。圖 2 說明立體視覺的裝置，LC 與 RC 分別代表左右兩相機的鏡頭位置。

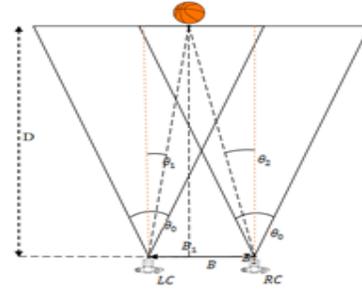


圖 2: 立體視覺裝置

在電腦視覺系統中，影像前處理是一個重要且常用的方法，影像前處理可以增強影像品質與改進運算效率。在抓取立體影像之後，將降低影像解析度來改進運算速度。例如，原解析度由 6000x4000 像素降為 600x400 像素。根據文獻所提 [14-17]，當解析度降至一確定等級並不影響系統的準確性。另一種改進運算速度的方法則是將影像從 RGB 色彩轉換為灰階等級。影像資訊可能會在灰階的色彩空間來遺失，不過對距離量測的精確度並沒有很明顯的影響。

相距固定距離的兩相機水平且平行放置，實驗開始先由右邊相機(RC)拍攝第一張影像，再由左邊相機(LC)拍攝第二張影像，如圖 3 所示，B 表示為兩相機之間的距離， $\theta_0$  為相機的水平視角，物件的位置距離 D 可以用底下的幾何運算來求得：距離 B 為  $B_1$  與  $B_2$  的和，因為  $B_1 = D \tan \theta_1$  與  $B_2 = D \tan \theta_2$ ，所以可得：

$$D = \frac{B}{\tan \theta_1 + \tan \theta_2} \quad (1)$$

根據 LC 視圖：

$$\frac{\tan \theta_1}{x_1} = \frac{\tan(\theta_0/2)}{x_0/2} \rightarrow \tan \theta_1 = \frac{2x_1 \tan(\theta_0/2)}{x_0} \quad (2)$$

根據 RC 視圖：

$$\frac{\tan \theta_2}{-x_2} = \frac{\tan(\theta_0/2)}{x_0/2} \rightarrow \tan \theta_2 = \frac{-2x_2 \tan(\theta_0/2)}{x_0} \quad (3)$$

由式(2)+式(3)：

$$\tan \theta_1 + \tan \theta_2 = \frac{2 \tan(\theta_0/2)(x_1 - x_2)}{x_0} \quad (4)$$

所以

$$D = \frac{B x_0}{2 \tan(\theta_0/2)(x_1 - x_2)} \quad (5)$$

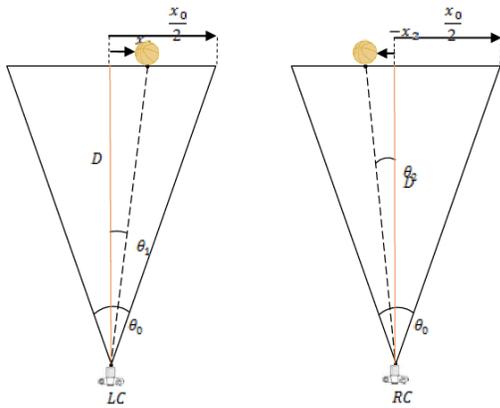


圖 3: 相機與物件的位置

其中 $x_0$ 為影像水平解析度， $x_1-x_2$ 為視差(disparity)。所以為了得到目標物件的距離 $D$ ，兩個相機所拍攝的影像需在同一個場景，兩張影像解析度降低為 $600 \times 400$ 像素，並由物件檢測器找出物件的位置。視差( $x_1-x_2$ )可由在目標物件座標的 $x$ 軸上的左相機與右相機的影像差來獲得。

### 2.3 機械手臂控制

本研究初步將利用深度學習影像辨識定位系統來控制六軸機械手臂，將雙目立體鏡頭水平安裝於手臂末端，如圖4所示，作為目標物件位置的影像辨識與影像深度的量測，使手臂能移動至目標物件的位置來驗證系統的可行性，其手臂影像定位處理流程如圖5所示：

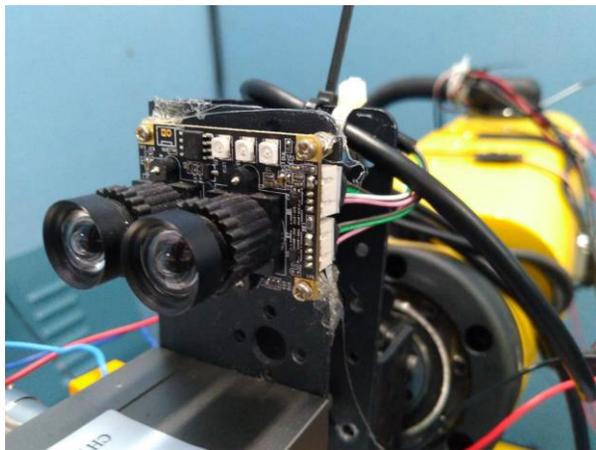


圖 4: 雙目立體視覺系統

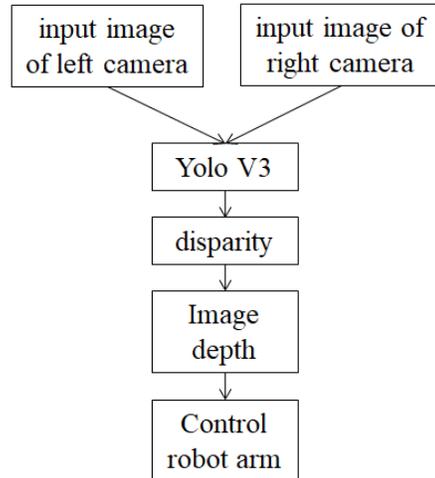


圖 5: 機械手臂影像定位處理流程

### 3. 結果與討論

本研究利用深度學習技術應用於乳牛乳房位置的辨識，採用 Yolo V3 演算法作為訓練時使用的模型，而用於模型訓練的乳牛乳房圖片有 200 張，50 張用於模型訓練後的測試。圖 6 為 Yolo V3 模型訓練過程的 Loss 變化情形，圖 7 為 Yolo V3 模型對不同的乳房進行測試的結果，系統能順利識別出乳房，及預測乳房位置，並使用邊框(bonding box)標記出乳房位置。

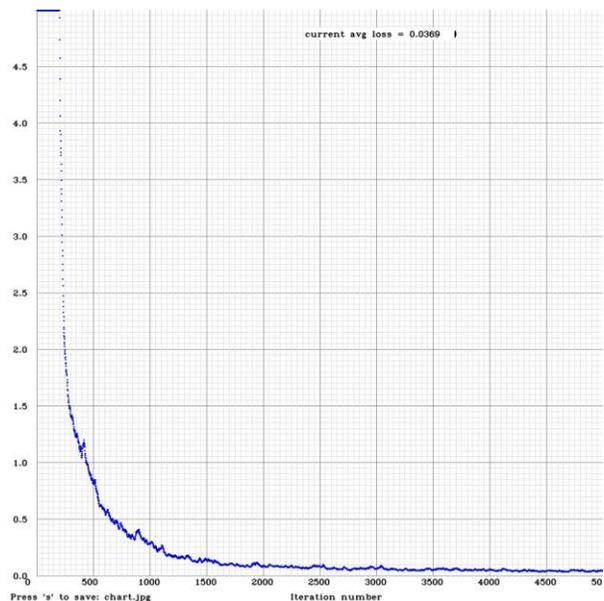


圖 6: Yolo V3 模型訓練的 Loss 變化情形

本研究以深度學習 Yolo V3 演算法的影像辨識技術進行目標物件的定位，透過立體視覺技術量測物件距離，再結合以 Borland C++ Builder 程式開發並完成使用者操作介面以及六軸機械手臂的控制程

式，使影像辨識、立體視覺測距以及機械手臂動作控制，完全在 Borland C++ Builder 環境下完成，其整合了機械手臂動作控制，以及深度學習立體視覺技術。

為了驗證本系統的可行性，本研究利用 yolo v3 影像辨識技術來定位乳頭模型的位置，計算出影像深度，再控制六軸機械手臂夾取榨乳杯並自動將乳杯套入乳頭模型，如圖 8 所示，首先完成以乳房模型進行神經網路訓練後，系統即可透過裝設於手臂末端的雙目攝影鏡頭擷取目標影像位置與計算出目標物件的影像深度，然後控制伺服馬達轉動。當手臂移動時，雙目鏡頭位置需始終保持水平，使系統可隨時追蹤目標物件，並使手臂能持續往目標位置移動。系統先拍攝兩張物件目標的左右兩張影像，然後由深度學習的技術來標定目標物件，藉由左右影像內的標定目標物件框的中心座標以求得視差，即可利用雙目視覺影像理論計算出目標物件的影像深度距離。

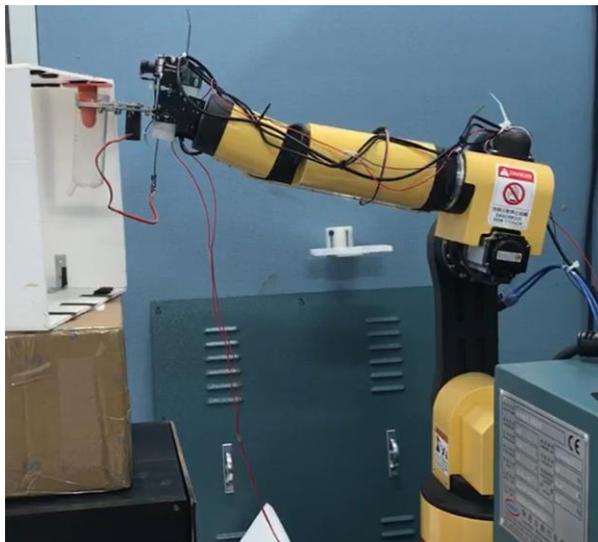


圖 8: 六軸機械手臂可將乳杯自動套入乳頭模型

本研究進行了不同物件深度距離的量測，每種物件深度重複量測 5 次，以求得平均影像深度與標準差(Std.)，影像深度量測結果如表一所示，由實驗結果發現，影像深度與實際深度的誤差隨著目標物越遠而越大，如圖 11 所示，對於雙眼視覺成像原理而言，目標物越遠，其像差值越小，使得影像深度距離的誤差變大。為了使機械手臂能精確的移動至目標位置，針對平均影像深度與誤差值進行曲線擬合(curve fitting)，找出其匹配的二次曲線方程式：

$$y=a+bx+cx^2 \quad (6)$$

其中  $y$  為實際距離與預測距離的誤差， $a=0.4536$ ， $b=-0.2119$ ， $c=0.1167$ ， $x$  為影像深度距離。由以上

誤差校正結果可以發現，機械手臂移動可透過二次曲線方程式(6)的誤差校正後，影像深度距離在 20cm 至 35cm 之間的範圍時，機械手臂可成功將乳杯自動套入乳頭模型。

#### 4. 結論

本研究利用 YOLO V3 影像辨識技術來定位乳頭模型的位置，以立體視覺技術計算出物件影像深度，再控制六軸機械手臂將榨乳杯自動套入乳頭模型，透過實驗來驗證系統的可行性，實驗結果發現，當物件距離越遠時，其影像深度的誤差越大，不過影像深度距離在 20cm 至 35cm 範圍時，機械手臂仍可成功將乳杯套入乳頭模型內。未來研究也將繼續改進影像深度的準確性，使系統更具實用性。

#### 5. 參考文獻

1. Yan, Z, Xiaodong, X, Xuejun, P, Wei, W (2003) Mobile robot indoor navigation using laser range finder and monocular vision. In IEEE International Conference on Intelligent Systems and Signal Processing, Faro, Portugal.
2. Andrea Cherubini, Fabien Spindler, François Chaumette, "A new tentacles-based technique for avoiding obstacles during visual navigation", Robotics and Automation (ICRA) 2012 IEEE International Conference on, pp. 4850-4855, 2012, ISSN 1050-4729.
3. Fengchun Dong, Sio-Hoi Ieng, Xavier Savatier, Ralph Etienne-Cummings, Ryad Benosman, Plenoptic cameras in real-time robotics, The International Journal of Robotics Research, vol. 32, 2: pp. 206-217
4. T. M. Mitchell, Machine Learning, vol. 45 (37), McGraw Hill, Burr Ridge, Ill, USA, 1997.
5. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, Cambridge, Mass, USA, 2016.
6. Adrian Carrio, Carlos Sampedro, Alejandro Rodriguez-Ramos, and Pascual Campoy, A Review of Deep Learning Methods and Applications for Unmanned Aerial Vehicles, Journal of Sensors, Volume 2017, Article ID 3296874, 13 pages
7. K. Fukushima, "Neocognitron: a hierarchical neural network capable of visual pattern recognition," Neural Networks, vol. 1, no. 2, pp. 119-130, 1988.
8. Y. Bengio, A. Courville, and P. Vincent, "Representation learning: a review and new

- perspectives,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 8, pp. 1798–1828, 2013.
9. J. Schmidhuber, “Deep learning in neural networks: an overview,” Neural Networks, vol. 61, pp. 85–117, 2015
  10. L. Tai and M. Liu, “Deep-learning in mobile robotics - from perception to control systems: a survey on why and why not,” CoRR abs/1612.07139. <http://arxiv.org/abs/1612.07139>.
  11. J. L. Sanchez-Lopez, M. Molina, H. Bayle et al., “A multilayered component-based approach for the development of aerial robotic systems: The aerostack framework,” Journal of Intelligent & Robotic Systems, pp. 1–27, 2017
  12. Data Science and Robots Blog. How do Convolutional Neural Networks work? [https://brohrer.github.io/how\\_convolutional\\_neural\\_networks\\_work.html](https://brohrer.github.io/how_convolutional_neural_networks_work.html)
  13. S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: towards real-time object detection with region proposal networks,” in Advances in Neural Information Processing Systems, vol. 28, pp. 91–99, 2015.
  14. Hsu, T.-S., & Wang, T.-C. (2015). An Improvement Stereo Vision Images Processing for Object Distance Measurement. International Journal of Automation and Smart Technology, 5(2), 85-90.
  15. Appiah, N., & Bandaru, N. (2011). Obstacle detection using stereo vision for self-driving cars.
  16. Mrovlje, J., & Vrancic, D. (2008). Distance measuring based on stereoscopic pictures. Paper presented at the 9th International PhD Workshop on Systems and Control: Young Generation Viewpoint.
  17. Kun Qian, Xudong Ma, Fang Fang, Xianzhong Dai, and Bo Zhou, Mobile robot self-localization in unstructured environments based on observation localizability estimation with low-cost laser range-finder and RGB-D sensors, International Journal of Advanced Robotic Systems, September-October 2016: 1–11