學　　生：曹雅筑
指導老師：徐百輝

以多視立體影像結合機器學習進行三維場景重建

一、 前言

1.1 研究背景與動機

在空間資訊領域，空間中的場景分析及目標物定位是重要課題，三維場景重建是現階段熱門的研究標的，透過物件與場景重建，以視覺化方式描述構造物或目標物並加以定位，輔助人類的觀察力，提升對空間中景物分布的了解，進而對該空間進行分析與研究，因此三維場景重建有其發展潛力與重要性。

迄今，三維場景重建的資料來源包括航空拍攝影像、衛星影像、光達(Light detection and ranging，LiDAR)點雲資料(Point cloud data)等，傳統多以衛星影像或航空拍攝取得資料，建立已知方位的立體像對，以人工逐點量測像點和解算，取得共軛像點，進而計算與產製數值地形模型(Digital terrain model，DTM)，藉以形塑三維地貌，後續隨科技發展，自動化的課題被推行，如影像相關(Image correlation)或影像匹配(Image matching)部分，以快速找到共軛像點，省去人力與時間成本，提升效率與自動化程度，使得攝影測量邁入資訊時代(曾義星，1997)，更與電腦視覺領域接軌，朝地圖或空間資訊數值化發展。

由多角度獲取的二維影像常被應用於三維場景重建，利用兩張以上不同視角的影像，模擬人類視覺系統，基於視差(Parallax)原理獲取影像對應點(Corresponding point)之間的位置關係，恢復出三維資訊，重建成本低、設備簡單、不接觸及不直接傷害場景，但傳統影像處理方法仰賴人工判釋與計算，若面對大量圖像資料，其處理成本會更加高昂，儘管電腦視覺與影像處理將以往人類處理範疇交由電腦執行，能在較短時間內完成被交付的作業，然後續為評估電腦處理的準確度與成功率，仍多仰賴人工介入處理，僅能被歸類為近半自動式處理程序。

機器學習(Machine learning)是目前新興使用的全自動式處理方法，仰賴程式語言與訓練模型，從過往的資料和經驗中學習，找到其運行規則或演算法，最後期望利用模型來做預測，更自我評估整個模型的成功率，解決原先須大量人工的作業項目。

現今三維建模廣泛應用在各領域，使用率增加，更追求其作業效率與精度，許多文獻與研究已針對小尺度物件與小範圍場景重建，在處理二維影像的領域展現了優秀的表現，然大尺度場景重建目前仍是挑戰，若能有效率且自動化地重建大尺度場景，將能對各領域有很大的幫助，如：虛擬實境(Augmented and virtual reality, AR and VR)(洪國隆，2007)、災區重現、室內外設計、機器人導航(Ann *et al.*, 2016)、三維虛擬智慧城市打造、古蹟重現等，貼近且活用於現實世界，因此本研究將提出一種基於多視立體影像，輔以機器學習直接從中重建三維模型的方法，簡化數據處理作業，使整個流程更具效益或達到程序全自動化。

二、 文獻回顧

本研究以三維場景重建全自動化為目標，利用多視立體影像，輔以先前相關研究，發展出一合適的三維機器學習演算法，進而完成三維建物模型重建，且顧及其精確度、真實度、細膩度，為此，本節將針對影像建模方式、機器學習框架、以多視立體影像進行三維建物模型重建等議題之相關研究作歸納整理。

2.1 影像建模(Image-Based Modeling)

二維平面影像與三維立體物件的差異在於深度的有無，視差、光彩、圖像使人類視覺系統可感知到多種深度線索(Depth cues)，進而形塑畫面幾何空間感，其中影像視角多寡與場景或物體的三維幾何形態將影響畫面空間感的重建，故本節將根據二維影像的選擇與其三維模型幾何形態呈現之相關研究作歸納整理。

2.1.1 單視角影像重建(Single View Reconstruction)

傳統攝影測量技術仰賴兩張以上影像進行立體測圖，使得單視角之影像較少被用於還原三維場景。若欲以一張像片建立深度圖，可利用影像色彩飽和度(Saturation)與物件紋理清晰度或銳利度之比較，簡易判斷其深度，如距離攝影機較近的物體，飽和度越大，紋理越清晰(Murata *et al.*, 1998)，藉以推斷物體與攝影機的距離關係，其中顏色特徵可利用影像中 RGB 與明度-飽和度-亮度(Hue-Saturation-Value，HSV)色彩坐標轉換，計算出飽和度數值；紋理特徵的部分可將彩色轉換成灰階影像，進而辨別場景類別。然目前尚無有效量化此兩數值分析，無法準確判斷場景中組合物的深度值，故該法並不可靠。

亦有研究提出利用場景邊緣特徵萃取，基於透視投影原理，偵測消失點(Vanishing point)及匯集於消失點的消失線(Vanishing line)，由畫面各像素與消失線之相對位置推求深度分布(Battiato *et al.*, 2004)。但該方法穩定性較低，不一定對各種場景皆有效，且像片涵蓋場景或物件的範圍少，易受障礙物遮蔽影響，且相機姿態過於單一，無其他姿態輔助補強幾何網形(莊曜誠，2013)，若要產製三維點雲，該點雲的數量也不夠覆蓋整個目標物，造成目標物模型有部分缺漏。

2.1.2 雙視角與多視立體重建(Multi-view Stereo Reconstruction)

雙視角影像模仿人類視覺，以雙角度觀察目標物，而多視立體(Multi-view stereo，MVS)透過多個不同視角位置取得影像資訊，解算相機內外方位參數，基於視差原理獲取影像對應點的位置關係，以影像特徵匹配演算法獲得特徵點，現今常用的影像匹配演算法如 Scale-invariant feature transform(David Lowe, 1999)，簡稱 SIFT，用於偵測與描述影像中的局部特性，在空間尺度中尋找極值點，在特徵提取的過程中，不易受到影像旋轉、縮放和灰階值差異影響，其後續也有基於 SIFT 進行改善之研究，如 SURF，提高解算過程的效率，應用範圍包含物體辨識、影像追蹤和動作比對、機器人地圖感知與導航、3D 模型建立等等。

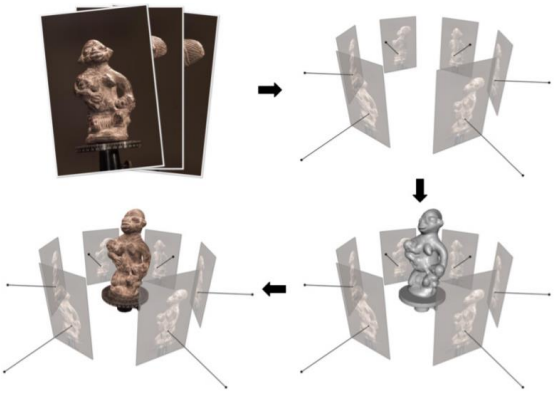


圖 1　多視角拍攝(Furukawa *et al.*, 2015)

實際操作上，可利用多台相機同時拍攝影像，或使用單一相機由多個視角取得影像進行重建，多台相機拍攝之優勢為可獲得豐富資訊來重建，藉以定位靜止物或對移動物進行軌跡追蹤，然相機設備的多寡直接影響研究與實驗成本，故成本考量下，單相機拍攝方法較為常見，目前應用單相機多視角拍攝的方法如運動回復結構(Structure from motion) (Snavely *et al.*, 2008)，利用多個場景自動恢復相機運動和場景結構，完成相機追蹤與影像匹配，然輸入像片越多，運算量越大，若序列影像相距太遠、基線距離較大，其重建效果將明顯降低(Ziegler *et al.*, 2003)。

過去 Brown 等人(2003)針對近代立體視覺與三維場景重構的方法做了更廣泛的整理，將重點放於圖像對應與匹配演算法(Correspondence methods)、如何解決場景障礙物的遮蔽問題(Methods for occlusion)與於現實中的即時(Real-time)應用(Brown *et al.*, 2003)，後續在 A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms(Seitz *et al.*, 2006)一文對多視立體影像重建三維模型的演算法作分類，提供具真實三維表面模型的影像資料庫，提及過去立體視覺演算法的優勢、缺陷以及後續如何追求改善，最後確立了評估三維重建成果的兩種量化指標：完整性(Completeness)與精度(Accuracy)，為其後 MVS 的發展樹立方向與指標，如 Accurate, Dense, and Robust Multi-View Stereopsis(Furukawa and Ponce, 2010)一文便以上文為基礎提出改良的方法，如對遮蔽區域資料缺失的改進、如何更自動化進行包圍體階層(Bounding volume hierarchies)的建立、剃除分類過程中的離群值(Outliers)與障礙物(Obstacles)等等，並強調該方法作業效率之提升，象徵了 MVS 發展的進步。

表 1　單視角與多視角影像比較

影像	單視角影像	多視角影像
取得方法	由單一視角進行拍攝	由多個不同視角進行拍攝
相機姿態	單一	多樣
全部像片涵蓋範圍	少	多
適用場景	少數特定場景	各式各樣場景
處理資料量與時間	少量省時	大量費時
模型精度與完整性	較低	較高

2.1.3 三維模型呈現(3D model representation)

A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms(Seitz *et al.*, 2006)闡述場景表現(Scene representation)的種類，場景表達是指使用特定數學模式進行表達三維立體場景，如體素(Voxel)、水平集(Level-sets)、多邊形網格(Polygon mesh)及深度圖(Depth map)等方式，除上述以外，多數文獻也提及點雲是一有效的表現方式，本節將針對較常見之三維模型表示進行整理。

(1) 三維像素(Voxel)

又名體積像素(Volume pixel)或體素，亦可稱為體積表示(Volumetric representation)，其與像素概念類似，像素為描述二維影像的單位，而體素則用於描述三維空間，如同三維網格(3D grid)應用於三維成像，實際應用如網路遊戲的虛擬場景呈現、醫學影像等領域，Photorealistic Scene Reconstruction by Voxel Coloring(Seitz and Ponce., 1999)一文便是以三維像素呈現其重建後的目標物，如下圖 2 所示，上方圖為灰階立體像素，下方圖輔以色彩呈現，越高三維解析度，其模型越貼近物件原始形狀，是多數神經網路常使用的三維模型表現方式。

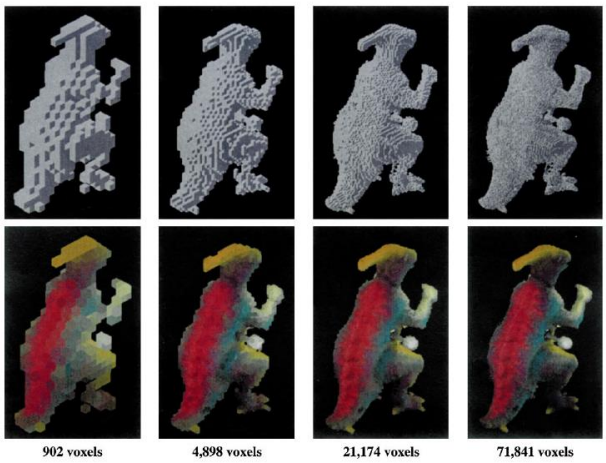


圖 2　以三維像素呈現三維模型(Seitz and Dyer, 1999)

(2) 多邊形網格(Polygon mesh)

首先產製三維點雲，再將表面點雲相連，進而表示成一組多邊形網，而這種多邊形多為三角形、四邊形或其他凸多邊形。在 A Practical Guide to Polygon Mesh Repairing(Campen *et al.*, 2012)一文講述多邊形網格是一常見的模型表現方式，簡介多邊形網格於各式各樣三維場景模型重建的應用，Real-time scene reconstruction and triangle mesh generation using multiple RGB-D cameras(Meerits *et al.*, 2017)一文更提及一新穎方法，其優勢為可利用多 RGB-D 深度感測器即時建立穩固的場景三角網格。

(3) 深度圖(Depth map)

透過視差可進行深度計算，形塑深度差異(Depth disparity)，呈現場景或物體的前後空間關係，目前主流的深度圖產生方法為利用單視角影像(Saxena *et al.*, 2008)及雙視角影像產生深度圖，多視角影像利用比對的方式求出兩張或多張畫

面的對應點，透過對應點之水平坐標差推導出深度圖，產製深度圖的品質較穩定且精確，但其計算複雜度更高，下圖 4 為一原始影像與其對應的深度圖(Scharstein and Szeliski, 2002)，透過深度圖呈現場景中的物件前後交錯關係，以影像匹配演算法找出遮蔽區(Occluded regions)或深度不連續區(Depth discontinuity regions)，結合匹配成本函數計算匹配成本值再行加總，最後優化視差圖並產出視差成果。

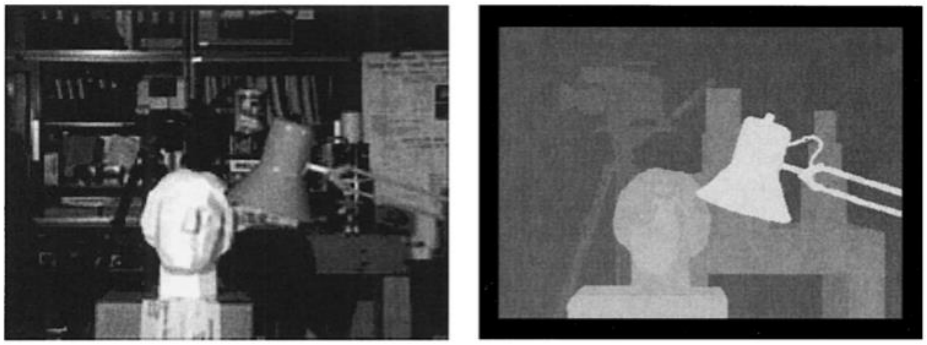


圖 4　深度圖之產出(Scharstein and Szeliski, 2002)

2.2 深度學習(Deep learning)

2.2.1 深度學習之概念

根據最早提倡人工智慧(Artificial intelligence)專家的說法，人工智慧為一具有智慧的機器，特別是電腦程式，讓電腦以近似人類的思考方式作業，機器學習為人工智慧的一環，Tom Mitchell 在其著作 Machine Learning(1990)一書提及，將大量資料交予機器直接進行訓練後，若機器具備學習能力，經過不斷執行的作業經驗，其產出的作業成果表現會更好，其中人工神經網路(Artificial neural networks)是機器學習的著名演算法之一，模仿生物神經系統(Biological nervous systems)的運作，由多層簡單神經元(Neurons)互相連結成網狀結構，結構內具有一組隨不斷學習而調變的權重或參數，輸入的資料經網路演算及統計後，輸出最佳化結果。

深度學習是機器學習的新分支，機器透過多個處理層(Layers)處理大量無序性的資料，學習完成特定工作，自動提取出特徵(Feature)以代表資料特性，取代過去萃取特徵所花費的時間，除影像特徵分類(Classification)及非線性迴歸(Non-linear regression)外，也可應用在維度調整、自然語言處理(Natural language processing)、人臉辨識等，甚至讓電腦具有自動產生語句及圖像的能力。

2.2.2 應用深度學習於二維影像重建三維模型

現階段已發展出的深度學習框架有卷積神經網路(Convolution neural network，CNN)、循環神經網路(Recurrent neural network，RNN)、Multi-view CNNs、3D-CNNs、3D-R2N2 與 SurfaceNet 等等，除處理歐基里德資料(Euclidean data)型態外，延伸至非歐基里德資料(Non-Euclidean data)，其中歐基里德數據的處理較常見，如下圖 4，對描述元(Descriptors)、投影(Projections)、RGB-D 深度資料、體積(Volumetric)資料及多視角影像，經過各自合適的演算框架，匯出最佳成果，在這些框架中許多框架奠基於監督式學習(Supervised learning)，在事先訓練

(Training)的過程中告訴機器答案，於資料上賦予標籤(Label)定義特徵，其中 CNNs 在影像識別方面的性能最為合適，故許多影像辨識模型多以 CNN 架構為基礎做延伸。

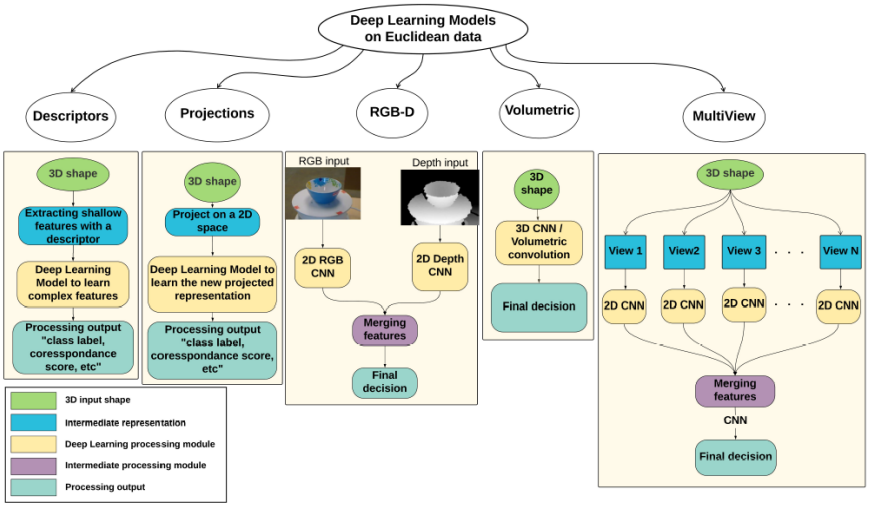


圖 4　對應不同三維資料的深度學習演算法(Ahmed *et al.*, 2018)

2.3 以多視立體影像之相關神經網路研究與挑戰

直接以多視立體影像發展的卷積神經網路 Multi-View CNNs(Su *et al.*, 2015)首先被發展於三維形狀辨識(3D shape recognition)，如圖 5，針對三維多邊形網格模型建立一系列二維多視角影像，匯入 CNNs 進行訓練，建立形狀描述元(Shape descriptor)，經池化(Pooling)後，再進行一次 CNNs 演算，最後推估影像為何物；Qi(2016)等人則以 Multi-View CNNs 為基礎，增加體積網格表示的三維模型及不同型態的多視角影像進行訓練，型態包含多解析度(Multi-resolution)、尺度(Scale)差異、不同方位角(Azimuth)等，加強了 Multi-View CNNs 的物件辨識能力。

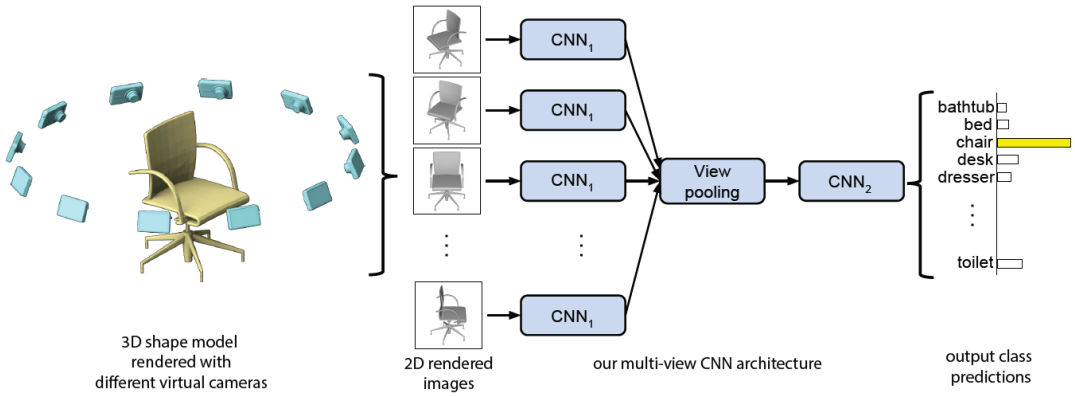


圖 5　Multi-View CNNs 之架構(Su *et al.*, 2015)

除物件辨識外，Learnt Stereo Machines(LSM)(Kar *et al.*, 2017)致力於用較少數量的多視角像片重建目標物，先匯入影像建立特徵圖，再反投影回三維特徵網格，利用 3D CNNs 作處理，產出體積像素圖或深度圖，文中適當引入已知相機姿態與先驗形狀，更好地回復影像間幾何位置關係，亦探討遮蔽區(Occlusion area)無法重建的幾何問題，針對遮蔽區加入更多影像進行預測與重建。

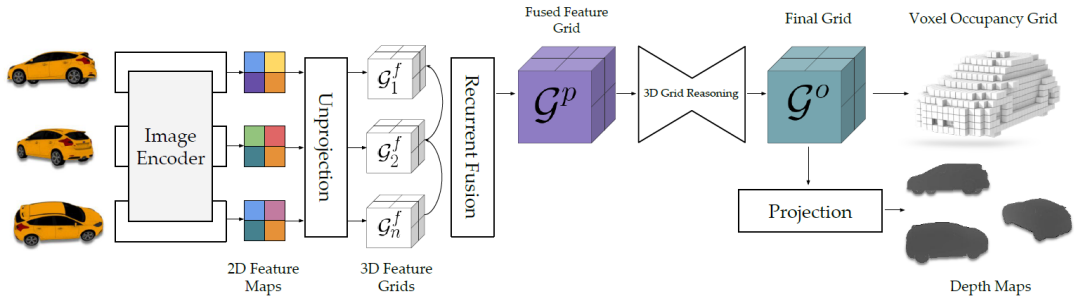


圖 6　Learnt Stereo Machines 之架構(Kar *et al.*, 2017)

綜觀多數文獻所言，重建三維模型時已發現些許問題，目前文獻中實驗目標多為小型單一物件，若要擴增至重建三維場景，有其難度；目標物遮蔽區影像不足，導致模型部分缺失，然為了遮蔽區而增加更多影像，將耗費時間與演算空間；若影像解析度不一，演算模型不一定能全然吸收且妥善處理，而高解析度的影像處理易造成記憶體不足或演算時間冗長；許多方法建議加入先驗資料，但先驗資料耗費人工處理，雖加入越多先驗資料，易使重建效果越加優良，但建立龐大資料庫需耗費大量成本；因此仍有需多待改進之處與將面對的挑戰。

三、 研究方法

本研究欲提出一個適合多視立體影像重建三維場景的深度學習演算法，以卷積神經網路(CNNs)為基礎架構，提升圖像特徵分類之準確性及效能，致力於改善以往發現的問題。

3.1 卷積神經網路

如圖 7 所示，卷積神經網路由一輸入層(Input layer)、許多隱藏層(Hidden layers)及最後輸出層(Output layer)所組成，隱藏層分成特徵偵測層(Feature detection layers)及全連結層(Fully connected layer)，其中特徵偵測層主要由卷積層(Convolutional layers)、池化層(Pooling layer)所組成，特徵偵測層可重複多次，得到不同層次的特徵，而全連結層(Fully connected layer, FC)則用來進行影像分類或辨識，指派前一層所偵測到的特徵至對應類別，本小節將概述此各部分的結構。

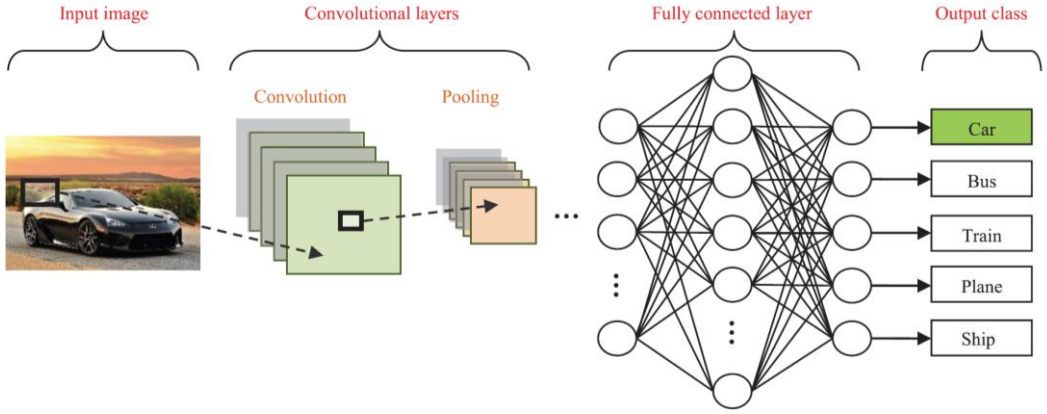


圖 7　卷積神經網路基本架構圖(Rawat and Wang, 2017)

3.1.1 卷積層(Convolutional layers)

在認識卷積層前，須先了解卷積層內的組成元素：濾波器(Filter)，又稱為特徵偵測器(Feature detector)或核心(Kernel)，是一 m×n 的二維矩陣，會在圖片上進行滑動，與圖中覆蓋到的像素進行運算，可自行設定滑動步伐(Stride)大小，提取圖片中重要的特徵，形成特徵圖(Feature map)，其中的計算便稱為卷積。

在進行卷積加權計算後，有可能會產生負值，常使用線性整流函數(Rectified Linear Unit, ReLU)去掉負值，正值維持原樣，亦即模擬神經細胞行為，當訊號量超過某個門檻值(Threshold)時，才允許訊號通過或輸出數值，是神經網絡中常用的一種激活函數(Activation function)，更能淬煉出物體形狀，作業效率更佳(Krizhevsky *et al.*, 2012)，隨滑動步伐越大，特徵圖越小，得視需要調整或擴增原始圖像邊緣(Padding)，不讓卷積後的圖像變小，故卷積運算可用在影像銳化、去除雜訊或提取感興趣的視覺特徵。

3.1.2 池化層(Pooling layer)

池化層以非線性降取樣(Downsampling)方式簡化卷積層的輸出，輸出值代表濾波器掃過範圍內的特徵統計值，根據計算出來的值不一樣，分為均值(Average)池化層與最大值(Maximum)池化層，象徵提取局部均值與最大值，一般常見最大值池化層，可自行設定濾波器大小、步伐等，調整網路所需訓練的參數數目。

3.1.3 全連接層(Fully connected layer)

全連接層旨在進行影像辨識與分類，將之前經過特徵偵測的結果平坦化(Flatten)，形成一維陣列，減少特徵位置對分類的影響，最後進入 *Softmax* 邏輯回歸函數(Softmax regression)，計算所有節點的輸出屬於各個類別的機率，將該節點歸類至最大機率對應的分類選項，其中 *Softmax* 回歸是邏輯回歸 (Logistic regression)的推廣，邏輯回歸適用於二元分類問題，而 *Softmax* 回歸適用於多元分類問題。

3.2 多視立體影像相關深度學習架構

目前深度學習的框架甚多，選擇適合的框架將能使作業事半功倍，本研究欲以端到端(End-to-end)深度學習網路 MVSNet(Yao *et al.*, 2018)作為研究框架，以此進行多視立體影像重建三維物件或場景，本節將 MVSNet 依功能分成四小節進行概述。

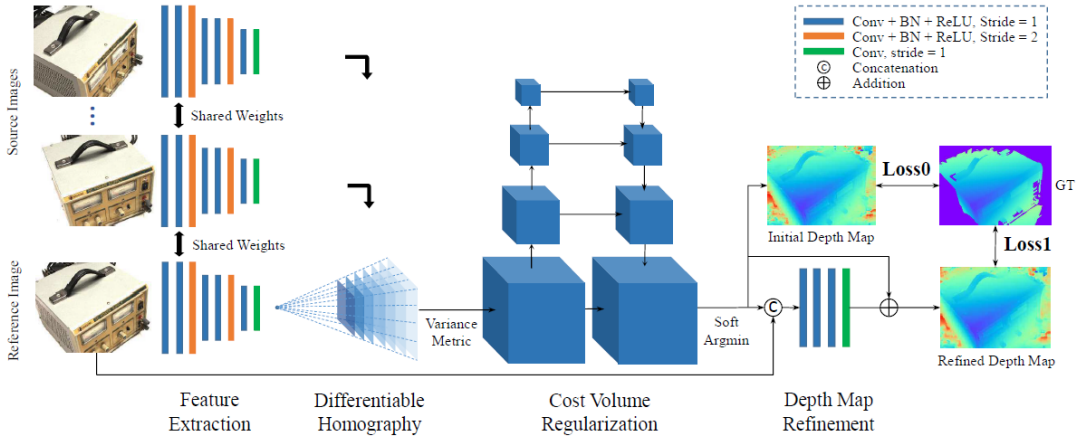


圖 8　MVSNet 基本架構圖(Yao *et al.*, 2018)

3.2.1　特徵萃取

輸入的多視角影像與其對應的已知相機幾何資訊（包含內方位參數、旋轉及平移參數）會經過二維特徵萃取神經網路，該網路由八個隱藏層構成，第三層與第六層的滑動步伐為 2，其餘滑動步伐為 1，另外，除第八層之外，每個隱藏層皆附有批次標準化(Batch-normalization layer)與線性激活函數 ReLU，將深度特徵圖分成三個層次，以不同等級表示特徵，形成特徵塔(Feature tower)，其中批次標準化是將資料一批一批逐一進行標準化計算，希望維持各個影像或特徵圖的尺度（平均值與標準差等），方便神經網路的訓練進行。

3.2.2　可微分之單應性矩陣轉換(Differentiable Homography)

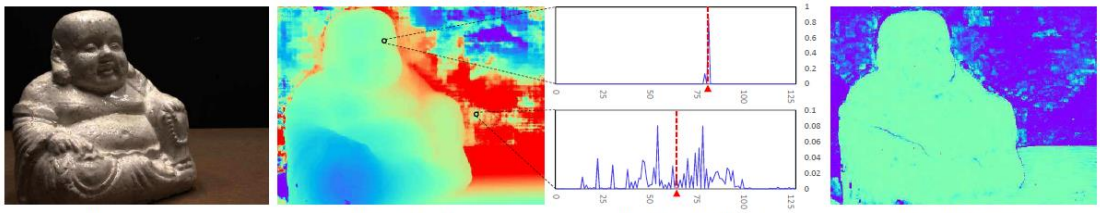
單應性矩陣是由基礎矩陣(Fundamental Matrix)或必要矩陣(Essential Matrix)計算而來，用於計算立體像對中原始影像上的某一像元對應至核線影像上像元間之轉換關係，完成核線影像重建(丁皓偉，2014)，由於輸入的多視角影像皆附有其已知相機幾何資訊，因此可將上一步驟產出的特徵圖進行單應性矩陣轉換，以便後續進行匹配成本值的計算。

3.2.3　匹配成本值計算與正規化(Cost Volume Regularization)

完成特徵圖的核線影像重建後，便可進行匹配成本值計算。在此依據 Yao(2018)提出的成本量測函數(Cost metric) M，用於計算多視角相似性量測(N-view Similarity measurement)，如下式(1)，透過輸入特徵圖(Input feature map)的長 H(Height)、寬 W(Width)、深度值 D(Depth sample number)與波段數 F(Channel number)定義特徵值 V(Feature Volume)的大小，而 M 函數用於計算所有特徵值的變異數，求得成本值 C (Cost volume)，如式(2)，再用成本值 C 推斷深度機率 P(Probability volume)，最後引入 multi-scale 3D CNNs 輔以 *Softmax* 函數執行機率正規化(Probability regularization)計算，去除深度圖中的雜訊與遮蔽區造成的影響，產製機率圖(Probability map)，其中 *Softmax* 為歸一化指數函數，經常用在神經網絡最後一層，作為輸出層，輸出 0 至 1 間的值作為樣本屬於各種類的機率表示，延伸進行多元分類。

$$V = \frac{W}{4} \cdot \frac{H}{4} \cdot D \cdot F \tag{1}$$

$$C = M(V_1, \cdots, V_N) = \frac{\sum_{i=1}^{N} (V_i - \overline{V_i})^2}{N} \tag{2}$$



(a) Reference image　(b) Inferred depth map　(c) Probability distribution　(d) Probability Map

圖 9　機率圖之產出範例(Yao et al., 2018)

3.2.4　深度圖精化(Refinement)與點雲模型產出

此部分將參考圖像(Reference image)及其初始深度圖(Initial depth map)將作為輸入資料匯入深度精化過程，深度精化過程包含兩部分，第一部分進行深度殘差學習網路(Depth residual learning network)運算，計算相同大小的參考圖和初始圖間的殘差，殘差包含正值與負值，不利用激活函數將負值消除，經計算完的殘差圖融合初始深度圖後作為精化深度圖輸出，產出圖像對應的三維點雲模型。其中在初始或精化深度圖的產出前，皆會引入真實深度圖(Ground truth depth map)，以 *Loss* 函數計算初始或精化深度圖與真實深度圖的差異，評斷神經網路的好壞。

3.3　用於重建三維模型的訓練資料

本研究希望先以線上開放的可用圖像訓練數據集進行測試，例如 ImageNet(Krizhevsky *et al.*, 2012)、DTU(Aanæs *et al.*, 2016)等，由小尺度物體進行著手，進而延伸至大場景重建，大場景之訓練資料則將取用現階段已經過處理的多視角無人機影像，以該影像先行產製的數值高程模型作為深度與三維模型重建的驗證資料，測試變因包含不同尺度、解析度大小、視角數量多寡、先驗資料匯入的多寡及物件或場景複雜程度等，並考慮與其他神經網路成果做比較，如 SurfaceNet、3D CNNs 等。

3.4　研究流程

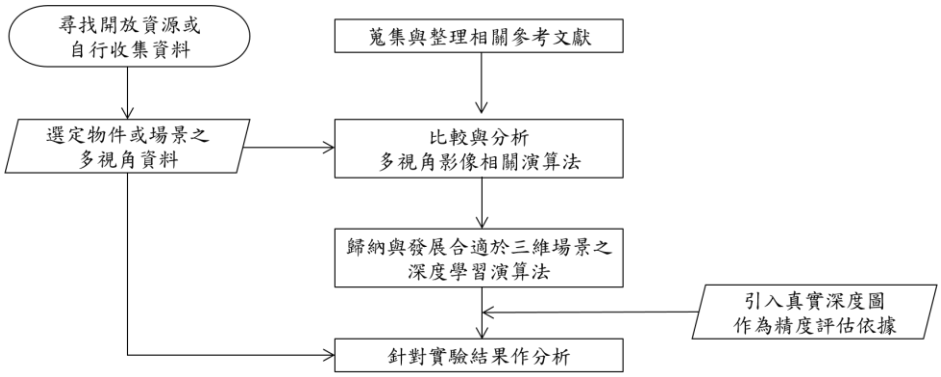


圖 10　研究流程圖

四、 預期成果與未來工作

本研究希望能根據蒐集整理之文獻以及比較各方法的結果與分析，發展出一種有效應用多視立體影像重建三維場景的方法，節省傳統方法所花費的時間及人力成本，簡化數據處理的工作流程，達到流程自動化之效，不失模型精度與完整性，亦預期將應用範圍延伸至大型場景重建。

後續本研究欲選定合適的建模標的物，完成相關文獻之蒐集及整理，比較與分析多視立體影像相關之深度學習演算法，如以 MVSNet 為基礎，進一步精進的 Recurrent-MVSNet(Yao *et al.*,2019)，最後歸納與發展適合於三維場景重建的深度學習演算法，並針對實驗成果作分析，同時反覆檢視與修正演算法。

五、 參考文獻

Aanæs, H., Jensen, R. R., Vogiatzis, G., Tola, E., and Dahl, Anders Bjorholm, 2016. Large-Scale Data for Multiple-View Stereopsis, *International Journal of Computer Vision*, *120*(2), 153-168.

Ann, N. Q., Achmad, M. S. H., Bayuaji, L., Daud, M. R., and Pebrianti, D., 2016. Study on 3D Scene Reconstruction in Robot Navigation using Stereo Vision, *2016 IEEE International Conference on Automatic Control and Intelligent Systems*, Selangor, Malaysia, pp. 72-77.

Battiato, S., Capra, A., Curti, S., and Cascia, M. L., 2004. 3D stereoscopic image pairs by depth-map generation, *2nd International Symposium on 3D Data Processing, Visualization and Transmission*, Thessaloniki, Greece, pp. 124-131.

Brown, M. Z., Burschka, D., and Hager, G. D., 2003. Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 25(8), 993-1008.

Campen, M., Attene, M., and Kobbelt, L., 2012. A Practical Guide to Polygon Mesh Repairing, Eurographics(Tutorials).

Furukawa, Y., Hernández, C. J. F., Graphics, T. i. C., & Vision, 2015. Multi-view stereo: A tutorial, *Foundations and Trends® in Computer Graphics and Vision*, 9(1-2), 1-148.

Furukawa, Y., and Ponce, J., 2010. Accurate, Dense, and Robust Multiview Stereopsis, *IEEE Transactions on Pattern Analysis and Machine Intelligence, 32*(8), 1362-1376.

Kar, A., Häne, C., and Malik, J., 2017. Learning a multi-view stereo machine, *Advances in neural information processing systems*, pp. 365-376.

Krizhevsky, A., Sutskever, I., and Hinton, G. E., 2012. Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, pp. 1097-1105.

Meerits, S., Nozick, V., and Saito, Hideo, 2017. Real-time scene reconstruction and triangle mesh generation using multiple RGB-D cameras, *Journal of Real-Time Image Processing*, pp. 1-13.

Meerits, S., Nozick, V., and Saito, Hideo, 2017. Real-time scene reconstruction and triangle mesh generation using multiple RGB-D cameras, *Journal of Real-Time Image Processing*, pp. 1-13.

Mitchell, T., Buchanan, B., DeJong, G., Dietterich, T., Rosenbloom, P., & Waibel, A., 1990. Machine learning, *Annual review of computer science*, 4(1), pp. 417-433.

Saxena, A., Chung, S. H., and Ng, Andrew., 2008. 3-D Depth Reconstruction from a Single Still Image, *International journal of computer vision*, 76(1), pp. 53-69.

Scharstein, D., & Szeliski, Richard, 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, *International journal of computer vision*, 47(1-3), pp. 7-42.

Seitz, S. M., Curless, B., Diebel, J., Scharstein, D., and Szeliski, R., 2006. A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms, *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 519-528.

Seitz, S. M., and Dyer, Charles. R., 1999. Photorealistic Scene Reconstruction by Voxel Coloring, *International Journal of Computer Vision*, 35(2), pp. 151-173.

Snavely, N., Seitz, S. M., and Szeliski, R., 2008. Modeling the World from Internet Photo Collections, *International Journal of Computer Vision*, 80(2), pp. 189-210.

Su, H., Maji, S., Kalogerakis, E., and Learned-Miller, E., 2015. Multi-view convolutional neural networks for 3d shape recognition, *2015 IEEE international conference on computer vision*, pp. 945-953.

Yao, Y., Luo, Z., Li, S., Fang, T., and Quan, L., 2018. MVSNet: Depth inference for unstructured multi-view stereo, *2018 European Conference on Computer Vision (ECCV)*, pp. 767-783.

Ziegler, R., Matusik, W., Pfister, H., and McMillan, L., 2003. 3D reconstruction using labeled image regions, *Eurographics Association*, Aachen, Germany, pp. 248-259.

丁皓偉，2014。結合十字區塊匹配之半全域匹配法優化作業，國立臺灣大學土木工程學研究所碩士論文，臺北市。

洪國隆，2007。使用立體視覺建立網路虛擬實境之地理資訊系統，國立臺灣大學生物產業機電工程學研究所碩士論文，臺北市。

莊曜誠，2013。基於三維建模與多視角影像擷取之物體三維定位及追蹤技術，國立中正大學電機工程研究所碩士論文，嘉義縣。

曾義星，1997。航空攝影測量如何邁向資訊時代，*航測及遙測學刊*，2(1): 103-112。