

基於 Petri Net 的 TSL 手勢即時辨識系統設計與驗證

The Real-Time Recognition System for Taiwan Sign Language Gestures Based on Petri Nets

孫子恩

林義楠*

明志科技大學電子工程系

明志科技大學電子工程系

摘要

台灣手語(Taiwan Sign Language, 簡稱 TSL)是台灣聽障社群採用之視覺溝通語言,其以手勢、表情及身體動作為要素,結合成一高度具備表達能力之語言系統。儘管, TSL 在溝通上扮演著重要角色,然對其了解與應用尚處相對有限的狀態。為克服這種溝通上的限制, TSL 識別系統應運而生。本文系統建置於樹莓派 4B Ubuntu20.04 平台上,運用影像識別技術 OpenCV 以及 Python 程式語言中的 Mediapipe 函式庫,可即時地辨識手勢,並將其結果顯示在螢幕上,以協助聽障人士與非聽障人士間的溝通。此 TSL 辨識系統亦透過基於數學理論和圖形特性的系統建模工具派翠網路(Petri Net),對其系統程式運作的架構,進行模型建構及驗證,以確保本文所提之系統架構之可行性、完整性、與可靠性。透過這些技術與方法,本研究有助於提升對 TSL 識別系統的理解與改善,以推動聽障人士與非聽障人士間溝通的效能與效果。在 SDGs 議題貢獻上,減少不平等(SDG 10)及對工業、創新與基礎設施(SDG 9)產生具體影響。

關鍵詞: 手勢辨識、樹莓派 4B、OpenCV、Python 程式語言、Mediapipe、派翠網路、SDGs

Abstract

Taiwan Sign Language (TSL) is a visual communication language used by the hearing-impaired community in Taiwan, combining hand gestures, facial expressions, and body movements into an expressive linguistic system. However, its practical application remains limited. To address this issue, this study proposes a real-time TSL gesture recognition system. The system is implemented on a Raspberry Pi 4B with Ubuntu 20.04, integrating computer vision techniques using OpenCV and MediaPipe for real-time gesture recognition. Recognition results are displayed instantly to facilitate communication between hearing-impaired and non-hearing-impaired individuals. In addition, Petri Net is employed to model and verify the system architecture and operational flow, ensuring its feasibility and reliability. This study enhances the development and application of TSL recognition systems, improving communication efficiency. It also contributes to the Sustainable Development Goals (SDGs), particularly Reduced Inequalities (SDG 10) and Industry, Innovation and Infrastructure (SDG 9).

Keywords: Taiwan Sign Language, Gesture Recognition, Raspberry Pi, Petri Net, SDGs

1. 前言

手語被視為一種視覺-空間語言，其以手部、身體和面部動作等方式來傳達意義，不依賴聲音及口語表達，而是透過特定的手勢、姿勢和表情以及空間中的動作及位置，以達到溝通和表達的目的。其主要功能在於彌合聽障人士與非聽障人士之間的交流障礙，提供一種有效的溝通工具。在台灣，TSL 被廣泛使用，被視為台灣語言和聽障人士群體的重要語言工具。並於 2018 年 12 月 25 日正式被列為台灣的國家語言。儘管 TSL 在拓展個體間差異方面具有顯著效益，但由於其複雜性對於非使用者而言學習難度較高，因此僅有寥寥數人對其有深入了解和掌握。為彌合聽障人士間的溝通障礙，TSL 識別系統應運而生。這種系統被設計為一種工具，旨在協助聽障人士之間更流暢地進行溝通，同時也為那些有意學習 TSL 的個體提供一個理解和學習的機會。此識別系統能夠作為聽障人士與非聽障人士之間的橋樑，進一步促進社會的多元和包容。藉由科技和教育的結合，為社會提供更多元的溝通管道，同時為社會共融與理解做出了重要貢獻。

以下介紹本文各節內容，第二節文獻探討。第三節說明系統架構。最後，第四節進行結論與未來發展方向討論。

2. 文獻探討

台灣手語(TSL)是台灣聽障人士所主要使用的手語形式，作為一種視覺語言，以手勢、手指、面部表情和身體動作等要素結合，被廣泛應用於交流和溝通之中。在台灣聽障人士中，台灣手語是他們主要的溝通工具，用於日常交流、思想和情感的表達、資訊分享以及社交連結的建立 [1]。本文關注台灣手語中的基本數字手勢表達，如圖 1 所示。這些數字手語詞彙被視為台灣手語體系中的核心元素。本文旨在對這些手語詞彙進行深入的學術研究，探討其結構、語義等。

OpenCV，即開放式電腦視覺庫(Open Source Computer Vision Library)，是一個具有開放原始碼的即時視訊處理工具，由 Intel 於 1999 年所開

發。其特點在於可供商業或非商業用途使用。OpenCV 涵蓋許多應用程式庫，包括臉部辨識、運動追蹤、機器學習以及移動機器人等 [2-3]。其廣泛的功能性和開放源碼的特性，使其在電腦視覺領域進行研究與開發時不可或缺的工具。

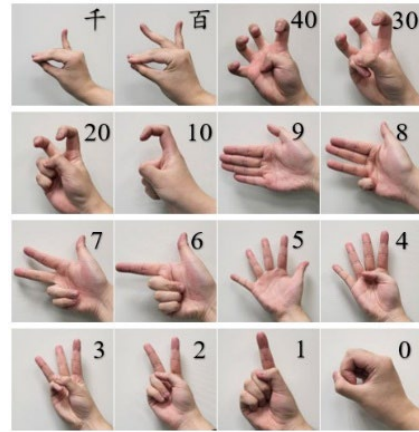


圖 1. 台灣手語-基礎數字手勢

Mediapipe 是 Google Research 所開發的開源軟體跨平台機器學習框架，其涵蓋多個平台程式語言，並能在 Web、Android、iOS、Linux、Windows、MacOS 以及邊緣設備上運行。其專注於即時處理視頻和音頻信號，以高效且準確的方式執行。Mediapipe 具備眾多功能，包括人臉追蹤、手掌偵測、姿勢偵測等 [4-5]。本文探索手勢即時辨識，著重選擇 Mediapipe 手掌偵測功能，透過這一技術，有效地追蹤手部動作，進行手勢辨別。運用手掌偵測技術，在視頻流中捕捉、辨識解釋手部動作，以滿足特定需求的即時準確手勢辨識。在運用 Mediapipe 手掌偵測技術時，系統會對偵測到的手掌進行精確解析，並標記其 20 個關鍵節點的位置，這些節點包含手指關節、手掌中心等，如圖 2 所示，以提供詳細而精確的定位信息 [5-6]。

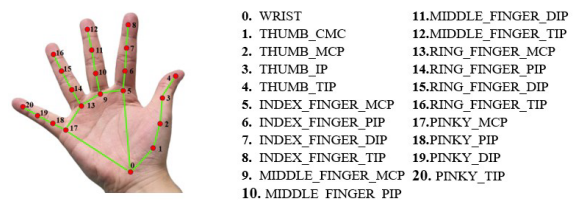


圖 2. 20 個關鍵節點位置

派翠網路(Petri Net)是一種結合數學理論與

圖形化表示的系統建模與分析工具[7][8][9][10]，由德國數學家 Carl Adam Petri 於 1960 年提出。其目的在於描述系統中事件、狀態以及流程間之動態關係，適用於並行、同步、資源共享與離散事件系統分析與模擬。Petri Net 以圖形化方式呈現系統行為，主要由三項基本元素構成，有位置(Place)：表示系統中的狀態或條件；轉移(Transition)：表示事件或動作；權杖(Token)：表示系統目前的狀態資訊。透過權杖在位置與轉移間流動，可模擬系統流程，進一步分析系統存在死結(Deadlock)、衝突或資源競爭等問題。

3. 系統架構

探討系統硬體架構及軟體流程圖。本文系統採用的硬體結構，包括元件配置、連接方式，以支持影像處理和模型驗證任務。接著，呈現軟體流程圖，描述系統開發過程中所使用的不同軟體工具和流程步驟，以確保系統的準確性和有效性。最終，闡述如何利用 WoPeD 工具進行 Petri Net 的建模和驗證，強調其在系統設計的有效評估方面的重要作用。透過本節的細緻解說，旨在展現系統開發過程中硬體、軟體和模型驗證的全面性和學術性，以支持系統的可靠性和有效性。

3.1 硬體架構

手語辨識系統的架構以樹莓派 4B 作為本地端主要平台，採用 Ubuntu 20.04 作為操作系統環境。系統的前端採用鏡頭作為主要影像輸入來源，而後端則是透過外接的顯示屏來呈現最終的辨識結果。在正常影像輸入的情況下，軟體將通過程式流程，將辨識完成的手語結果顯示在外接的顯示屏上。該系統的基本架構符合一般嵌入式系統的設計理念，藉由樹莓派 4B 提供的計算能力與 Ubuntu 20.04 穩定的操作環境，透過鏡頭和外接顯示屏的配合，實現從影像輸入到最終辨識結果輸出的完整系統流程，如圖 3 所示，可清晰呈現系統基本架構和各部分之連接關係。



圖 3. 系統示意圖

3.2 軟體流程

本研究系統軟體主要分為 4 段程式碼，涵蓋了 3 個函數(vector_2d_angle、hand_angle、hand_pos) 及 1 個主程式功能：vector_2d_angle()，為計算角度的函數是一種數學計算方法，其主要目的是獲取兩個給定座標之間的角度值。其流程基於兩個二維向量 v1 和 v2，利用向量的數學性質和計算方法來求解兩向量之間的夾角。

當鏡頭啟動並捕捉使用者手部時，手語辨識系統會擷取手部的 20 個關鍵點。例如，對於大拇指的辨識，該系統依賴於大拇指和手掌四個關鍵節點的配置，將它們視為兩個線段的端點，以計算手指的彎曲程度。這一計算過程涉及以下數學公式：假設 $v1_x = x2 - x1$ ($v1$: x 座標), $v1_y = y2 - y1$ ($v1$: y 座標)；同樣地， $v2_x = x4 - x3$ ($v2$: x 座標), $v2_y = y4 - y3$ ($v2$: y 座標)，將其轉換成兩個向量 v1 和 v2。根據向量公式可得：

$$\cos \theta = \frac{v1 \cdot v2}{|v1| \cdot |v2|} \quad (1)$$

若以座標表示向量，則 $v1 = (v1_x, v1_y)$ 和 $v2 = (v2_x, v2_y)$ ，公式為：

$$\cos \theta = \frac{(v1_x \cdot v1_y + v1_y \cdot v2_y)}{(\sqrt{v1_x^2 + v1_y^2}) \cdot (\sqrt{v2_x^2 + v2_y^2})} \quad (2)$$

$$\theta = \cos^{-1} \frac{(v1_x \cdot v1_y + v1_y \cdot v2_y)}{(\sqrt{v1_x^2 + v1_y^2}) \cdot (\sqrt{v2_x^2 + v2_y^2})} \quad (3)$$

hand_angle()：手指角度函數，旨在根據傳入的 20 個節點座標，計算並返回手指的角度值。其流程透過利用計算角度的函數，計算兩個向量之間的夾角來實現手指角度的推斷。

hand_pos()：返回手勢名稱函數，是基於手指角度的串列內容，用於識別手部手勢並返回對應的手勢名稱。其流程將手指角度串列中的角度值分配給各自變數，再根據手指角度進行手勢辨識並返回對應名稱，當沒有滿足任何特定手勢的條件時，返回空字串。這種手勢識別方法是基於手指角度的數值條件，通過特定角度組合辨識手勢。

主程式 main()，運用 Mediapipe 框架中手部

偵測模型，對手勢進行識別。主要步驟涵蓋鏡頭捕獲即時影像、將影像尺寸調整以提升運算效率、轉換影像色彩空間至 RGB 格式。接著利用 Mediapipe 辨識手部關鍵特徵，計算手指角度並擷取與手勢相關資訊。最後在影像上加入文字標示識別到的手勢，透過 OpenCV 視窗呈現處理後之影像。

3.3 Petri Net 建模與驗證

本文系統流程健全的系統驗證與運作可行性運用 Petri Net 對系統流程進行建模驗證，確保辨識流程之完整可行性。本系統，依 WoPeD 建模，具 23 個 Places、29 個 Transitions，及 58 個 Arcs。建模驗證，證實系統符合 Petri Net 規範，具完整性及無死結可行性，如圖 4、圖 5 所示。

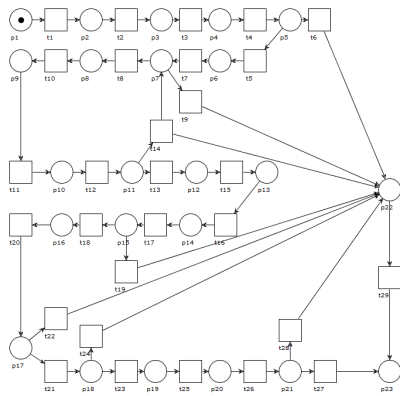


圖 4. Petri Net 建模

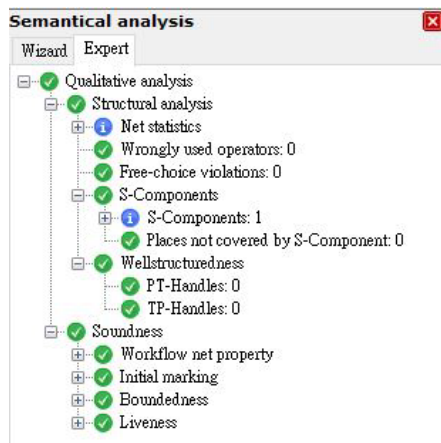


圖 5. Petri Net 驗證結果

關於具體的位置點和轉移點的詳細說明，本文在 Petri Net 模型中的流程為：起始於 P1(系統流程準備開始)，接著進行 T1(開始)操作。之後，

進入 P2(判別環境是否有攝像頭)並經由 T2(判別成功)，接著 P3(載入手部偵測模型)，透過 T3(載入成功)，接著進入 P4(獲取手部偵測模型的參數)，再經由 T4(獲取成功)，進而進入 P5(鏡頭是否開啟)。若鏡頭成功開啟 T5(Yes)，則繼續進入 P6(設置影像尺寸)；反之，若鏡頭開啟失敗 T6(No)，則系統進入 P22(輸出錯誤字串)。判別 T7(設置成功)後，進入 P7(影格是否有讀取到)。若影格成功讀取 T8(Yes)，則繼續進入 P8(獲取影格並調整為指定大小)；若影格讀取失敗 T9(No)，系統進入 P22(輸出錯誤字串)。在 T10(獲取及調整成功)的判別後，進入 P9(讀取影格轉換為 RGB 格式)，判別 T11(轉換成功)，隨後進入 P10(載入 hand process 函數)。經過 T12(載入成功)的判別，系統進入 P11(手部偵測是否成功)。若手部偵測成功 T13(Yes)，則進入 P12(建立 finger_point 列表)；反之，若偵測失敗 T14(No)，P7(重新讀取)。在獲得 T15(獲得 finger_point 空列表)，進入 P13(將 20 個節點座標記錄到 finger_point 列表)若紀錄成功進 T16(紀錄成功)，接著進入 P14(載入 hand_angle 副程式)，經過 T17(載入成功)的判別後，進入 P15(hand_angle 是否執行成功)。若執行成功 T18(Yes)，則進入 P16(載入 hand_pose 副程式)；若執行失敗 T19(No)，則進入 P22(輸出錯誤字串)。在 T20(載入成功)的判別後，進入 P17(hand_pose 是否執行成功)。若執行成功 T21(Yes)，則進入 P18(載入 OpenCV 顯示中文函式庫是否成功)；若執行失敗 T22(No)，則進入 P22(輸出錯誤字串)。若載入成功 T23(Yes)，則進入 P19(調整文字位置大小顏色)；若載入失敗 T24(No)，則進入 P22(輸出錯誤字串)。在 T25(調整成功)判別後，進入 P20(將結果顯示在 OpenCV 視窗)，並獲得 T26(獲得結果)，接著，系統進入 P21(判斷是否按下鍵盤 q 鍵)，若是 T27(Yes)，則進入 P23(系統流程結束)；否則 T28(No)，返回到 P7(影格是否有讀取到)重新執行。所有的 P22(輸出錯誤字串)會進入 T29(結束)，最終進入 P23(系統流程結束)。

3.4 系統運行結果

本研究於光線明亮且背景單純之環境下進行系統測試，受測者以手心向外方式示範預先定義之 16 種手勢，其手勢示範如圖 6 所示。本測試建立數字手勢標準，作為後續評估手語辨識系統正確性之依據。



圖 6. 辨識結果

本文共定義 16 種靜態手勢，並針對每一手勢進行 10 次辨識試驗，整體辨識準確率達 95.6%。如圖 7 所示。圖中以手勢名稱為 X 軸，辨識正確次數為 Y 軸。實驗結果顯示，本系統於多數手勢辨識上皆具有良好表現。然而，在手勢十、二十、三十與四十支辨識過程中，由於需判斷小指彎曲接近 90 度之細微差異，較容易與手勢一、二、三與四產生混淆。此外，在未經額外優化之情況下，系統在 Raspberry Pi 4 Model B 搭配 Ubuntu 20.04 環境中，平均處理速度為 12-18 FPS，已達基本即時互動需求。系統採逐幀影像處理方式，對每一影格進行手勢辨識，因此當使用者連續變換手勢時，系統可即時更新辨識結果。

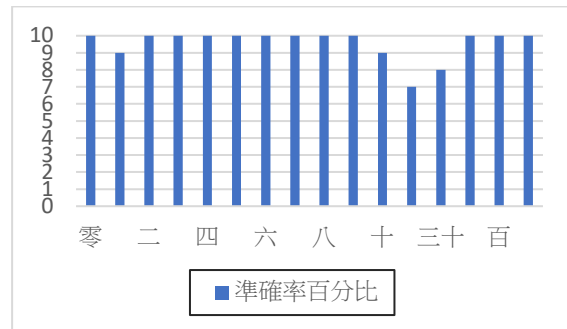


圖 7. 測試準確率百分比圖

3.5 成果評估指標

為全面評估所提出的手語辨識系統之有效性和信賴性，本文採用精度(Precision)、召回率(Recall)、正確率(Accuracy)和 F1 Score 值等指標來評估模型性能，實驗成果評估指標值，如表 1 所示。精度(Precision)：是指模型在將樣本預測為正類別時，實際上為正類別樣本所佔之比例，為輸出結果的可信度，其計算式為： $\frac{TP}{(TP + FP)}$ 。其中，

TP(True Positive)：即正確手勢預測為正確手勢。而 FP(False Positive)：即錯誤手勢預測為正確手勢。本文計算為 153 除以(153 加上 4)，結果為 0.974。召回率(Recall)：指系統偵測能力性能，即模型能正確檢測出所有真實正類別樣本之比例。其計算公式為： $\frac{TP}{(TP + FN)}$ 。其中，FN(False Negative)：未

能正確辨別的正類別，即正確手勢預測為錯誤手勢。本文召回率為 153 除以(153 加上 2)，結果為 0.987。正確率(Accuracy)：模型對整個測試樣本集中的樣本進行正確預測的比例，用以衡量模型在整體測試數據上的準確性表現，即系統判斷準確度，代表整體可靠性，計算式為： $\frac{TP+TN}{Total}$ 其中，TN(True Negative)：模型正確地將負類別預測為負類別，即錯誤手勢預測為錯誤手勢。Total：總測試次數。本文正確率為(153 加上 1)除以 160，結果為 0.962。F1 Score：是一綜合指標，用於評估模型在分類性能。結合模型精度及召回率，視為是精度和召回率的調和平均值，提供一種平衡度量，考慮模型在正類別和負類別樣本中之表現。F1 Score 值：同時考慮誤判與漏判，值域[0, 1]，越接近 1 表示又準又完整，而越接近 0 則表示性能越差，計算公式為： $2 * \frac{P * R}{(P+R)}$ 。其中 P：精度，R：召回率。

本文精度(0.974)和召回率(0.987)相乘後除以兩者的總合乘以 2，得到 F1 Score 值(0.98)。

表 1. 實驗結果評估指標

Precision	Recall	Accuracy	F1 Score
97.4%	98.7%	96.2%	0.98

4. 結論

本文實現一基於樹莓派 4B 與 Ubuntu 20.04 手語辨識系統，以鏡頭作為視覺輸入裝置，透過系統實現地端影像處理與手語辨識，並呈現辨識結果於外接螢幕，系統之 F1 Score 值達 0.98 之 TSL 手勢即時辨識系統。實現對於聽障社群溝通需求的回應，透過鏡頭捕獲即時影像，系統能夠快速、準確地辨識手語，將結果呈現在螢幕，為聽障群族提供了一個便利實用的交流方式。在可持續發展目標(SDGs)貢獻，可減少不平等(SDG 10)及對工業、創新與基礎設施(SDG 9)等有具體影響。

5. 參考文獻

1. I-Tun Tseng, "A Study on the Development of Tactile Sign Language in Taiwan," 特殊教育發展期刊 65 期 pp. 23-37, 2018.
2. Jie Huang, Anmin Huang, Liming Wang,

- "Intelligent Video Surveillance of Tourist Attractions Based on Virtual Reality Technology," IEEE Access, Vol.8, August 2020.
3. Elias Kougiianos, Saraju P. Mohanty, Gavin Coelho; Umar Albalawi, Prabha Sundaravadivel, "Design of a High-Performance System for Secure Image Communication in the Internet of Things," IEEE Access, Vol.4, March 2016.
4. B. Natarajan, E. Rajalakshmi, R. Elakkiya, Ketan Kotecha, Ajith Abraham, Lubna Abdelkareim Gabralla, V. Subramaniaswamy, "Development of an End-to-End Deep Learning Framework for Sign Language Recognition, Translation, and Video Generation," IEEE Access, Vol.10, September 2022.
5. Tomas Simon, Hanbyul Joo, Iain Matthews, Yaser Sheikh, "Hand Keypoint Detection in Single Images using Multiview Bootstrapping," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.
6. Itsaso Rodríguez-Moreno, José María Martínez-Otzeta, Izaro Goienetxea, Igor Rodriguez, Basilio Sierra, "A New Approach for Video Action Recognition: CSP-Based Filtering for Video to Image Transformation," IEEE Access, Vol.9, October 2021.
7. Yi-Nan Lin, Sheng-Kuan Wang, Gwo-Jen Chiou, Cheng-Ying Yang, Victor R. L. Shen, Zhi Yang Su, "Development and Verification of an IoT-enabled Air Quality Monitoring System Based on Petri Nets," Wireless Personal Communications volume 131, pages63–87, 2023.
8. Cheng-Ying Yang, Yi-Nan Lin, Sheng-Kuan Wang, Victor R.L. Shen, Yi-Chih Tung, Jia-Fu Lin, "An Edge Computing System for Fast Image Recognition Based on Convolutional Neural Network and Petri Net Model," Multimedia Tools and Applications, 2023.
9. Cheng-Ying Yang, Yi-Nan Lin, Victor R.L. Shen, Frank H.C. Shen, Chien-Chi Wang, "A Novel IoT-Enabled System for Real-Time Face Mask Recognition Based on Petri Nets," IEEE Internet of Things Journal, September 2023.
10. Cheng-Ying Yang, Yi-Nan Lin, Victor R. L. Shen, Frank H. C. Shen, and Wun-Siang Jheng, "A Novel IoT-Enabled System for Real-Time Monitoring Home Appliances Using Petri Nets," in IEEE Canadian Journal of Electrical and Computer Engineering, vol. 48, no. 3, pp. 204-215, 2025.