

國立清華大學 電機工程學系

實作專題研究摘要

Airborne-Ultrasound Targeting Based
Human Positioning and Body
Orientation Detection

基於空播超音波標靶之人體定位

與人體朝向偵測

專題領域：系統組

組 別：B600

指導教授：李夢麟教授

組員姓名：111061130沈牧群、111061232廖桓毅、111061244田仁泰

研究期間：2025年2月1日 至 2025年11月24日止，共 10 個月

第一章、報告摘要

在智慧裝置與人機互動技術快速發展的背景下，精確的使用者定位與方向感知技術成為重要研究課題。本研究的核心在於透過超音波陣列技術，提升定位與方向感知的精準度，藉此偵測人體位置、距離及臉部朝向，使其能夠應用於筆記型電腦、智慧音響、醫療照護與駕駛監測等場景。

本研究旨在以低成本的空播超音波技術實現人體定位與朝向辨識。架構包含 HC-SR04 超音波模組、CUSA-TR80 超音波換能器與 STM32F407VET6 微處理機，微處理機採用單一 ADC (Analog to Digital Converter) 多通道輪流取樣並搭配 DMA (Direct Memory Access) 完成五通道高效率回波擷取。為提升資料品質，後端進行帶通濾波、Hilbert 包絡分析、線性插值修正與感測器靈敏度補償，使回波訊號更穩定且適合後續分析。

在定位方面，我們以 Receive beamforming 將各角度的回波延遲對齊後生成 B-mode 影像，再以影像亮度峰值估測人體位置。最終以 CNN (Convolutional Neural Network) 對 B-mode 影像進行分類，在 1Tx (Transmitter)、5Rx (Receiver) 的配置下可達 83% 定位準確率。

在朝向辨識方面，我們將 B-mode 影像作為輸入，以 CNN 進行姿態分類（面朝中間、面朝右、面朝左），辨識準確率達 86%，顯示影像化後的空間資訊具有足夠特徵以搭配機器學習模型判斷朝向。

整體而言，本研究證明低成本超音波模組經由適當的訊號處理與影像化，可成功實現人體定位與朝向偵測。未來可透過提升 Tx/Rx 陣列規模、改善波束成型與擴增資料集，以提升於複雜環境下的定位與辨識效能。

第二章、報告內容

2.1 研究動機與目的

隨著智慧設備與人機互動技術的發展，準確且高效的使用者定位技術變得愈加重要。現今市場上的定位技術主要可分為影像識別、藍牙、Ultra-wideband (UWB) 定位與超音波測距三類。影像識別雖然提供高解析度資訊，卻存在隱私問題、功耗高且在低光環境下效果不佳；藍牙與 UWB 雖具備低功耗特性，但定位精度受環境干擾較大，且難以達到毫米級準確度；而超音波測距技術因其隱私性、低功耗、高精準度、低光適應性的特點，仍有獨特市場價值。

如今，超音波定位技術已被廣泛應用在醫學、聲學顯微鏡等領域，但應用於人體定位的技術卻相對甚少。因此，本研究將專注於 airborne-ultrasound targeting 技術，透過超音波陣列波束聚焦，實現對使用者距離及朝向的偵測，並進一步探索其應用於空間音訊等領域的可能性。

2.2 研究方法

2.2.1 系統架構

本章將詳細說明系統如何實現人體定位與朝向辨識。整體系統可分為四個主要架構：

1. 硬體端建立
2. 前端訊號處理
3. Receive beamforming 與 B-mode 影像生成
4. 機器學習 (Machine Learning) 進行分類與結果量化

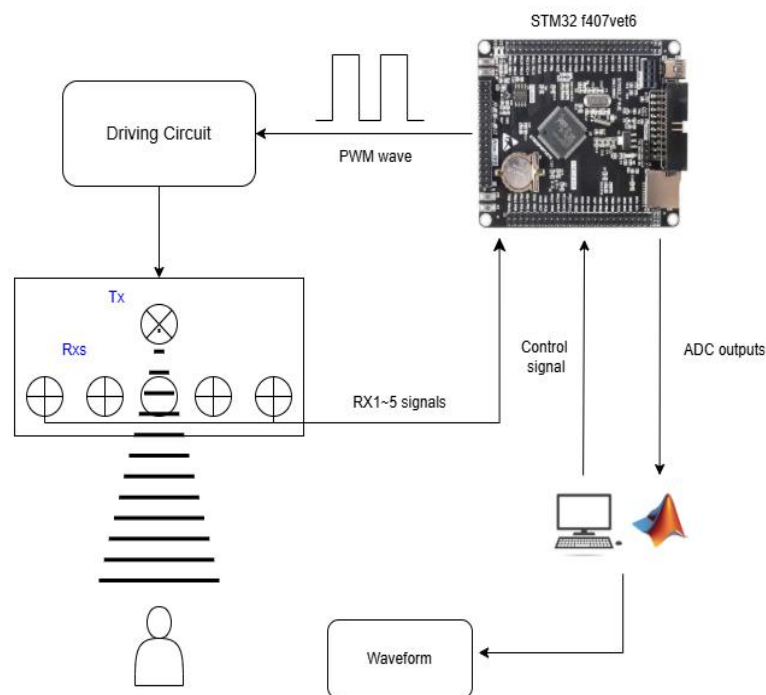


圖 1：系統架構圖

● 硬體端建立

硬體的部分主要涉及以下：

1. STM32F407VET6 開發板

本系統以 STM32F407VET6 微控制器為核心，負責產生 Tx 的 PWM 驅動訊號，並透過內建的 ADC 模組進行回波訊號的擷取。

為實現多通道回波的高效率擷取，本研究採用單一 ADC 多通道輪流取樣，並搭配 DMA (Direct Memory Access) 傳輸機制，有效完成多通道的高速資料擷取

2. 超音波陣列配置

本研究採用單一發射端進行訊號發射，接收端為五個 Rx 所組成的陣列，如下圖所示：

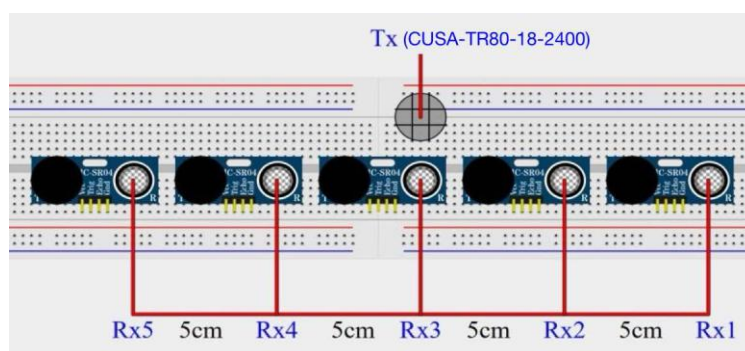


圖 2：Tx-Rx 陣列配置

本研究之發射端 (Tx) 採用 CUSA-TR80-18-2400 型號。由於原先使用的 HC-SR04 作為發射端時，其指向角僅約 $\pm 15^\circ$ ，規格無法滿足需求，因此改選用具備較廣指向角的 CUSA-TR80。

3. 超音波換能器驅動電路

為了驅動 CUSA-TR80 超音波換能器，需要提供週期性 Trigger 訊號以產生發射波形。在本研究中，我們透過 STM32 輸出 3.3V 的 PWM 方波，並利用 MOSFET 作為開關元件，將 PWM 連接至 Gate 端，將低電壓 PWM 轉換為高電壓方波來做驅動。

● 後端訊號處理

為提升回波訊號的品質，訊號在送入分析階段前會經過一系列優化處理：

1. 頻帶濾波

去除原始訊號中的雜訊，僅保留 40kHz 附近的有效資訊。

2. 飽和修正

針對 ADC 飽和的訊號區段，採用線性插值進行修正，使波形更接近原始真實值，進一步提升訊號品質。

3. Hann Window

取各通道訊號峰值為中心，加入時長 1ms 的 Hann Window，讓訊號變化更平滑。

4. 包絡分析

為了更清楚地觀察與分析訊號特徵，使用 Hilbert Transform 將原始波形轉換為包絡訊號，利於後續能量與時序的分析。

5. 正規化

我們將各通道的訊號進行正規化處理(Normalization)，使其落在統一範圍，以便比較與後續分析。

6. 元件靈敏度校正

由於不同 HC-SR04 感測元件存在靈敏度差異，實驗中先測量各元件間的能量偏差，再對各通道訊號加上適當的偏移量，以確保資料分析的一致性與準確性。

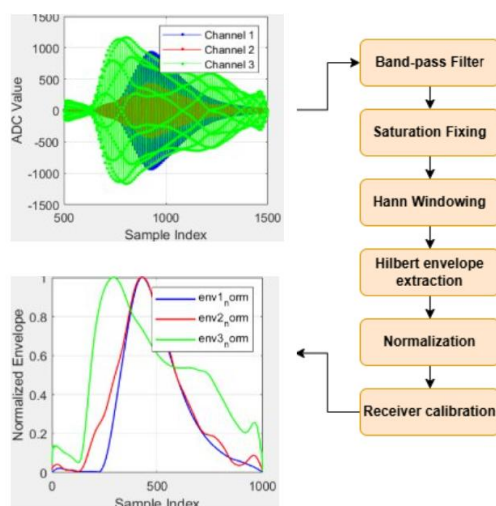


圖 3：訊號處理流程圖

● Receive beamforming 與 B-mode 影像生成

我們採用單一發射端進行訊號發射，總共 5 個接收端 (Rx) 會依據其物理位置計算對應的延遲時間做 Beamforming。後端系統則根據計算的延遲，將不同通道的接收訊號進行時間對齊與疊加 (Delay and Sum)，以形成一張 B-mode 影像的灰階圖，呈現目標區域的結構特徵。

我們目標是透過 Receive beamforming 實現對特定範圍的掃瞄，為達成此目的，我們可事先計算針對特定角度聚焦所需的延遲時間，並依此進行訊號疊加。計算方法如下：

$$\Delta t = \frac{s}{v} * \sin \theta$$

Δt 為 Rx 所需的延遲開啟時間、s 為 Rx 的距離、v 為聲速，在室溫下約為 343m/s、 θ 為和陣列平面法線的夾角，也是 Rx 聚焦的方向。

透過上述方式，我們分別為五個發射器建立了延遲時間陣列，控制 Rx 對 Tx

正負 35 度每 5 度做一次掃描，每一次掃描可獲得該方向上不同距離的能量分布。我們總共有 15 種角度，根據每一筆資料的延遲時間，給予時間上的 shift 後疊加，建構出一以 Tx 為中心、正負 35 度的扇形 B-mode 影像。

● 機器學習 (Machine Learning) 進行分類與結果量化

在本研究中我們主要使用 CNN(Convolutional neural network)模型，在人體定位以及朝向辨識。

Module 輸入： B-mode image

Module 輸出： Class label(面向正面、面向右方、面向左方)

Model structure： 2-layer Conv1D feature extractor + Dense classifier with dropout

Dataset： 我們每一種類別，皆取 400 筆當作 training set，100 筆當作 validation set，100 筆 testing set

2.2.2 人體定位

本研究 Beamforming 總共有 15 種角度，每種角度都對應一種延遲時間，且都有一筆距離-能量的資料，我們根據每一筆資料的延遲時間，給予時間上的 shift，讓每筆資料對齊，建構出以 Tx 為中心、正負 35 度的扇形 B-mode 影像。

在 B-mode 影像中，像素亮度取決於該位置的回波能量大小。能量越強，影像中該處越明亮。而影像中出現的亮點即代表物體可能的區域。透過搜尋影像中能量最大的像素位置，我們可以推斷該點為目標物體的位置，但我們發現圖中除了主亮帶，附近也有一微弱的亮帶，這是由 side lobes 產生的偽影。

2.2.3 朝向辨識

本研究所生成的 B-mode 影像範圍，是根據實際發射端 (Tx) 位置以及設定的掃描範圍繪製，因此可以將影像視為一個俯視角度的呈現。在影像中，亮帶區域代表回波能量較強的位置，即物體所在的區域。

然而，亮帶的形狀與分布會受到物體大小與距離影響。例如，若目標是一個傾斜 45° 的斜面，則影像中亮帶的一側會偏向後方，而另一側則偏向前方。透過觀察這種亮度分布的趨勢，我們可以推測物體的輪廓，並以此作為姿態判斷的依據。

2.3 研究結果

2.3.1 人體定位

在感測器的掃描範圍內，將觀測物置於多個不同位置並進行量測，如下圖所示。對每個位置量測 100 筆座標資料，去除極端數據後取其平均值，之後將量測結果與理論值比對以觀察其誤差。本實驗共使用兩種觀測物，分別為水壺與人體。下圖中，紅色叉叉代表物體實際位置，藍色叉叉代表測量出的位置，編號將各組實際值與測量值一一對應。

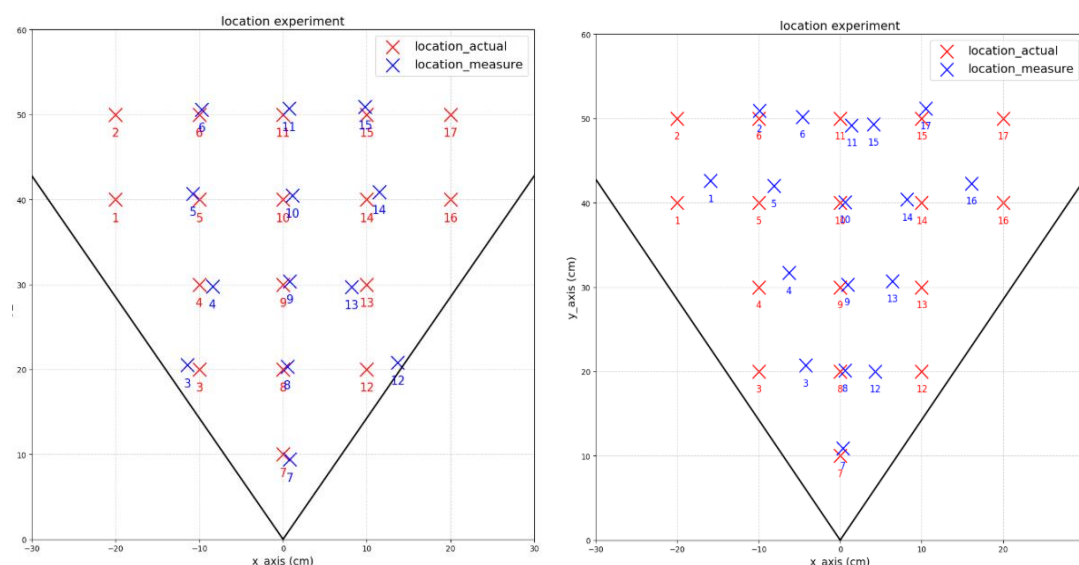


圖 4：水壺定位(左)&人體定位(右)實驗結果

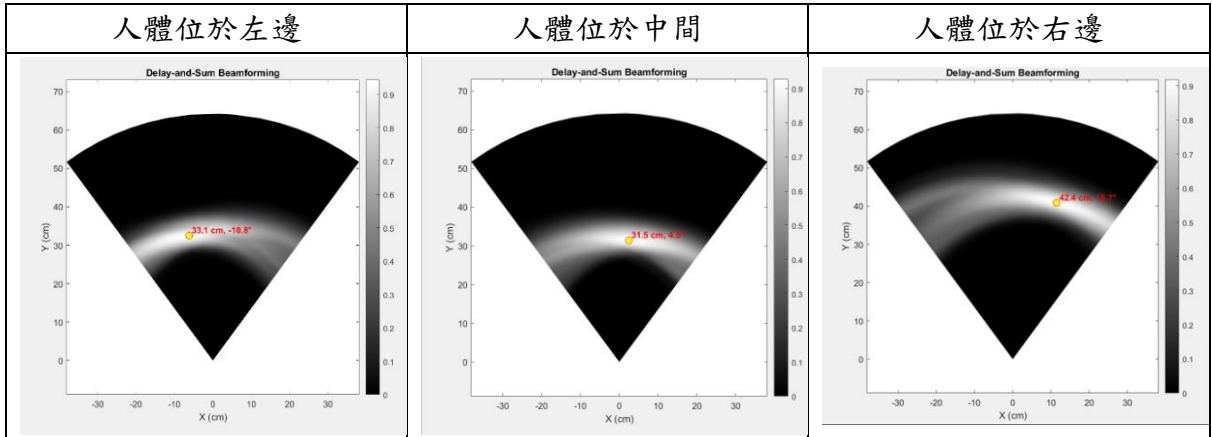
在水壺定位實驗中，測量位置與實際位置基本上差異不大，並且可以注意在 1、2、17、18 號位置由於位置跳動幅度過大，導致測量位置缺乏參考價值。這是因為 1、2、17、18 號位置與 Tx 的相對角度較大且距離較遠，導致 Rx 接收到的水壺反射能量很小，進而使 sidelobe 對 waveform 的影響變為顯著，因而導致了水壺的位置無法被準確的判斷。不過，此實驗說明了在小角度或近距離內的定點辨識仍有一定的準確性

在人體定位實驗中，測量位置對比實際位置有向內收縮的趨勢。這是因為人體是一個較大面積的反射面(人體的肩寬約 30cm 左右)。以 1 號位置舉例，來自人體左半部的反射回波能量較右半部強(人面對發射器)，因此測量位置便傾向人體面朝向的左邊(也就是圖像的中心)靠攏，並且這個偏差的現象會隨著距離越遠、角度變大而越明顯。雖然整體位置偏差較大，但 y 方向的距離基本正確，且顯示人體位置在左、中、右的趨勢是正確的，這已足夠我們進行之後的機器學習。

● 人體定位機器學習

在位置辨識測試中，我們將人體作為辨識目標，並將位置分為三類：中間(±10cm 以內)、左邊、以及右邊。

表 4：人體不同位置 B-mode 測試結果



實驗中，我們設計人體站於不同距離，並將每個方位（中間、左邊、右邊）收集 400 筆影像作為訓練集、100 筆作為驗證集，以及 100 筆作為測試集，用於計算分類正確率。

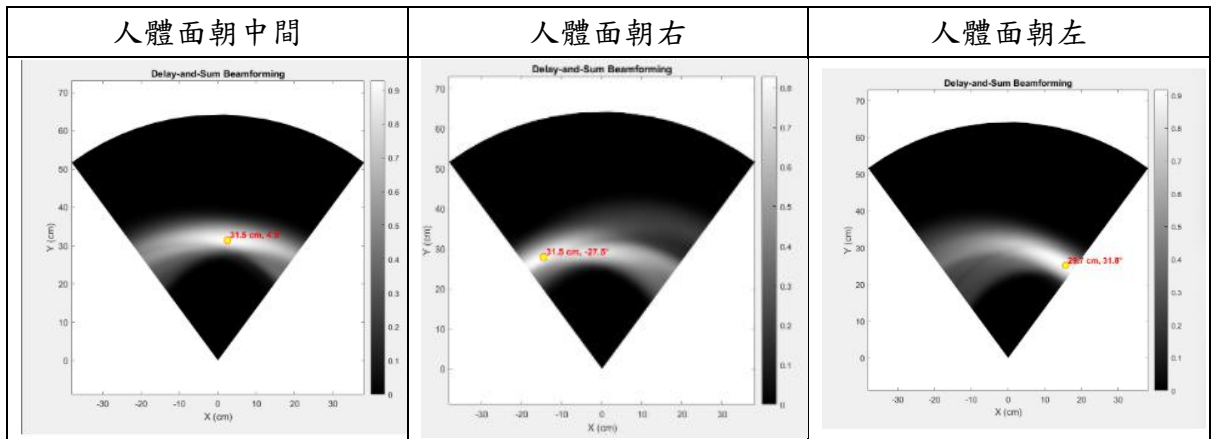
表 5：人體定位機器學習結果

配置	使用模型	正確率	結果分析
1Tx 5Rx	CNN	83%	由於人體會小幅度的晃動，在晃動時波型會變的不穩定，並且人體位於極端位置時(如靠近掃描邊緣)波型也會變得較不穩定，因而導致了正確率的下降。然而整體正確率有 80% 以上，說明我們的 B-mode 影像仍能基本呈現出人體位置的資訊。

2.3.2 人體朝向

在朝向辨識測試中，我們將人體作為辨識目標，並將朝向分為三類：面朝中間（旋轉角度在 ±10 度以內）、面朝左以及面朝右。

表 6：人體不同姿態 B-mode 測試結果



實驗中，人體距發射器 32 cm，每個位置與朝向收集 400 筆影像作為訓練集、100 筆作為驗證集、100 筆作為測試集，用於計算分類正確率。

表 7：人體姿態機器學習結果

配置	使用模型	正確率	結果分析
1Tx 5Rx	CNN	86%	使用 B-mode 影像訓練 CNN 後，辨識正確率顯著提升。相較於觀察感測器的回波波形，B-mode 影像能更清楚地呈現可辨識的特徵，對分類器而言更容易提取有效資訊。

2.3.3 結合機器學習技術判斷人體朝向與定位

在使用機器學習判斷人體定位與朝向後，我們打算將兩個特徵同時帶入機器學習訓練，以同時判斷人體的定位與朝向，定位與朝向的分類如下表九種(綠框代表人體位置)：

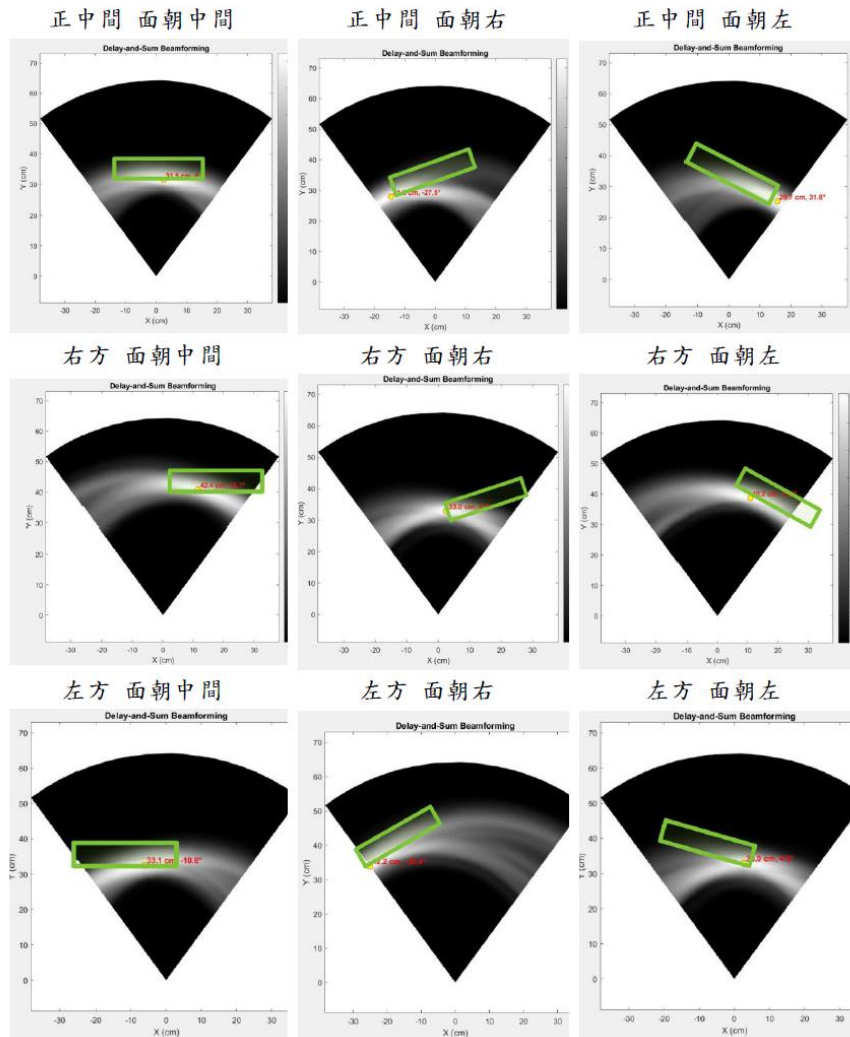


圖 5：人體朝向與定位機器學習結果

實驗中，人體距發射器固定為 32 cm，每個位置與朝向收集 400 筆影像作為訓練集、100 筆作為驗證集，以及 100 筆作為測試集，用於計算分類正確率。

表 9：機器學習模型預測結果

位置 & 朝向	正中 間面 朝 中間	正中 間面 朝右	正中 間面 朝左	右方 面朝 中間	右方 面朝 右	右方 面朝 左	左方 面朝 中間	左方 面朝 右	左方 面朝 左
正確率	17%	25%	14%	22%	25%	31%	21%	29%	14%

根據測試結果，本模型的平均辨識正確率僅約 22%，整體效能不如預期。推測主因在於部分類別的 B-mode 影像過於相似，使模型難以有效區分

雖然模型訓練結果未達預期，但影像亮帶的趨勢與我們的預測基本相符，未來可考慮改進硬軟體配置，包括增加接收陣列數量、提高發射能量、採用更大角度的 Tx/Rx 配置以減少偽影，並結合更進階的機器學習架構與更大規模的資料集，以改善辨識準確度並提升整體系統性能

2.4 總結

本研究以低成本的 HC-SR04 超音波模組與 STM32F407VET6 平台，成功建構一套具備多通道回波擷取、清晰 B-mode 影像生成，以及人體定位／朝向辨識功能的系統。透過單一 ADC 多通道輪流取樣、DMA 傳輸與多階段訊號前處理（帶通濾波、Hilbert 包絡、飽和插值與靈敏度補償），我們有效提升了回波訊號的穩定度，為後續影像化與機器學習分析奠定基礎。

在定位方面，我們使用 Receive beamforming 技術生成 B-mode 影像，使系統能在較大角度範圍內取得更具參考性的空間資訊。最終以 CNN 分析 B-mode 影像，可達到 83% 的定位準確率，證實影像化能有效減少傳統回波波型受角度與反射面的不穩定影響。

在朝向辨識方面，我們以 CNN 分析 B-mode 影像，可達到 86% 的朝向辨識準確率。結果顯示，B-mode 可清楚呈現人體朝向造成的亮度分布差異，且其空間資訊具有足夠特徵以搭配機器學習模型判斷朝向。

雖然整合定位與朝向的九類分類任務仍受偽影與影像相似度影響而表現有限，但整體研究已成功驗證：即便使用一般低成本超音波模組，只要搭配適當的訊號處理與影像化方法，即可實現具可行性的人體位置與朝向偵測。

未來可進一步提升效能的方向包含：

1. 增加接收器數量以改善波束成型解析度。
2. 採用指向角更大或能量更強的發射器以拓寬有效感測範圍。
3. 改善陣列配置與降低偽影產生以提升影像穩定性。
4. 導入更進階的深度學習架構（如 attention-based models）以強化複雜姿態的分類能力。
5. 擴充資料集並導入動態測試，邁向即時人體追蹤。

綜上，本研究證明超音波仍具作為低功耗、低成本且具隱私性的定位技術之潛力，並為空播超音波應用於人機互動與空間感知領域提供初步的實作驗證。

第三章、心得感想

我們的專題是李夢麟教授新開設的題目，由於題目聚焦於硬體與空中超音波，並非李教授擅長的領域（教授主要專注於生醫超音波與影像處理），因此大部分的技术仍須由我們自行摸索，因此實際準備起來比預期困難許多。然而，比起沿著前人已鋪好的道路完成一個「保險」的專題，我們更享受親手開創一條新的路徑，在不斷摸索與試驗的過程中獲得經驗與收穫。

整個專題最困難之處在於我們必須整合非常多來自不同領域的模組與技術，包括：

- **硬體端**：探頭選擇、驅動電路、訊號擷取與前級放大
- **微處理機端**：Arduino、STM32 的 ADC、DMA 與 PWM 控制
- **軟體端**：訊號處理、資料傳輸，以及基於 SVM 與 CNN 的特徵判斷

這些技術彼此跨越電子、通訊、DSP 與機器學習領域，讓我們在整合時不斷遭遇挑戰，例如驅動電壓不足、發射波形不穩定、接收雜訊過大、程式時序同步困難等，但也正因為克服這些問題，我們對跨領域研究有了更深刻的體會。

目前網路上關於「利用定點發射與接收器進行超音波辨識」的研究文獻仍非常稀少，而現今的電腦與手機大多是依靠紅外線或影像技術來進行定位與辨識，這更加突顯了本專題的新穎性與獨創性。我相信本專題的成果能成為後續研究的養分，並期待未來有一天能真正應用於產業，推動超音波在人機互動或智慧裝置中的新發展。