

## 具抗運動干擾之水上物件定位影像測距系統

黃仁傑\* 曾俊舜\*<sup>1</sup> 陳文忠\* 涂宗萍\*\*\* 黃添福\*\*\* 林政均\*\* 陳品涵\*

\*財團法人船舶暨海洋產業研發中心

\*\*國立臺灣海洋大學 電機工程學系、人工智慧研究中心

\*\*\*科力航太股份有限公司

<sup>1</sup>E-mail: cstsenglynn@soic.org.tw

關鍵詞：人工智慧、電腦視覺、水面物件偵測、影像測距、非穩態平台、抗運動干擾

### 摘要

相比 AI (artificial intelligence) 電腦視覺在陸上自駕車、無人空拍機等成熟應用，海洋船舶領域則存在諸多待開發的 AI 應用潛力，而 AI 電腦視覺應用於偵測辨識海面物件逐漸有團隊投入研究，其主要利用 AI 電腦視覺進行海上船舶避碰，期達到海上感知避碰自駕之目的[1][2]；順應這個趨勢，船舶暨海洋產業研發中心(SOIC)與國立臺灣海洋大學人工智慧研究中心(NTOU-AIRC)自 2019 年起合作開發 AI 電腦視覺技術應用於水上載具的相關應用，自 2021 年起更增加與科力航太股份有限公司合作致力於 AI 電腦視覺產品落地化。

不同於陸上環境，為能補償海上波浪、強風等造成水上載具的搖晃運動，本文提出一具抗運動干擾之水上物件定位影像測距系統，其以河上與海上場景為例，利用熱成像鏡頭搭配 AI 物件偵測器鎖定水面物件(船隻、浮標)，模型 mAP 92.21%，結合演算法估測目標距離，接著利用搖晃運動的規律性與週期性，將影像每幀測距結果以時序列執行抗運動干擾之濾波，以獲得抗運動干擾的水上物件最終距離；據此，大幅降低晃動所造成的影像測距誤差(±10%以內)。

### 前言

對於船舶的航行而言，航行路徑上可能遭遇的水上物件非常多樣。舉例來說，如浮標、浮木等障礙物，及其他行駛中的船舶等水上物件，均屬於船舶航行時需要注意的標的。

在有人駕駛船舶的狀況下，通常多半依賴駕駛的視覺判斷水上物件是否具碰撞的風險；然而，隨著智慧自主航行船舶的技術逐步演進，目前的船舶已經逐漸導入功能越來越強的船舶輔助駕駛系統(ship driving assistance system)。舉例而言，如透過 UC-win/Road 實現多種 3D/VR 空間表現的船舶操縱模擬器[3]已經被用於綜合船舶駕駛支援系統的開發與使用，船舶操縱模擬器(Ship handling simulator)可用模擬任意場域的情境，完整的航行模擬數據可用於訓練適合實際場域的 AI。另一方面，利用人工智慧 (artificial intelligence, AI) 的辨識技術以半自動甚至全自動駕駛已然為指日可待之事。例如勞斯萊斯 (Rolls-Royce) 和芬蘭國有渡輪運營商芬費里斯 (Finferries) 在芬蘭圖爾庫市南部的群島上展示了世界上第一艘全自動渡輪[4]，其所述全自動渡輪不僅具備自航能力之外，還可以運行簡單條件之自主避碰。在亞洲部份，韓國現代重工也開發現代智

慧導航輔助系統 (Hyundai Intelligent Navigation Assistant System, HINAS) [5]。上述案例均大幅推進船舶自主航行的商業化。

因此，利用影像辨識技術判斷水上物件的距離，使用如雙視覺影像、自動識別系統 (Automatic Identification System, AIS)、導航雷達 (Navigation Radar)、光達 (LiDAR) 或人工智慧，屬於現階段技術開發非常重要的一環。然而，實際上船舶航行時會遭遇到非常多的因素干擾，進而影響人工智慧的影像辨識判讀能力，包含：天氣條件、水面的反射、波浪的遮蔽、水流造成的船舶運動和姿態變化過於劇烈等環境因素干擾，均有可能影響到辨識距離的精確度，有鑑於此，如何善用人工智慧對於水上物件與攝像鏡頭之間的距離判讀能力，並且還能減少如前述環境因素所帶來的判讀誤差，為本文主要欲解決的問題。

### 文獻回顧

近年 AI 應用飛速發展，使幾乎所有領域均可見觸及 AI 演算法提升效能的實例，其中最著名的屬深度神經網路 (Deep Neural Network, DNN) 與電腦視覺 (Computer Vision) 結合應用於物件偵測 (object detec-

tion)，透過累積大量自然影像資料進行學習訓練，不斷地提升物件的偵測及辨識精確率。

物件偵測是一種與影像處理和電腦視覺相關的技術，用於檢測影像或影片中特定類別或表徵的對象，例如：人、汽車、建築物、貓、狗、船隻、浮標等，物件偵測能應用在許多領域，例如：監控系統[6]、醫療影像[7]、人臉辨識[8]，目前多數通用之物件偵測器能在影像中對物件定位且給予一個矩形邊界框(bounding box)，並提供辨識該物件的分類類別以及給予信心機率值，而近幾年來物件偵測模型不斷推陳出新，辨識準確率也大幅提升，辨識速度也日益精進。著名的物件偵測模型有：R-CNN、Fast R-CNN[9]、Faster R-CNN[10]、Mask R-CNN[11]、YOLO3[12]、YOLO4[13]、YOLO5[14]、SSD[15]、EfficientDet[16]，而這些物件偵測模型的核心演算法都是基於卷積神經網路(Convolutional Neural Networks, CNN)[17]進行各種改良而來。

### 研究方法步驟

針對本文方法步驟，系統架構如圖 1，詳細說明如下：

#### 1. 硬體設備準備與測試

##### 1.1. 熱成像鏡頭

圖 1 之「SFCV001」(圖 2)具備輔助避碰警示功能與「FLIR M364C」(圖 3)具備偵蒐功能，均為協力廠家科力航太提供之鏡頭；其中「SFCV001」為我們團隊特別客製，由三顆 640 x 512, 69° FOV, 9mm, 30Hz 組合而成。

##### 1.2. 計算單元 (compute unit)

本系統中最關鍵屬計算單元，如圖 4，係 BOXER-6840-CFL 高效能工業電腦，該計算單元有多組 I/O 介面及支援多螢幕輸出，並支援擴充高效能圖形顯示顯示卡輔助 AI 運算，其運算結果將傳輸顯示於多功能顯示器 MFD(Multi Function Display)。

##### 1.3. 硬體接線說明

據圖 1 所示，偵蒐與輔助避碰警示功能共計四組鏡頭，利用 BNC/AV 端子介面與 DVR 連接，藉 DVR 讓四組鏡頭畫面自動監視與錄影儲存，並藉 DVR 類比影像訊號轉為 RTSP 串流模式傳輸，轉變成類似 IP Cam 的傳輸方式，此設計的好處是計算單元只需與 DVR 共存於同一區域網路即可透過 DVR 的 IP 取得四組鏡頭的 RTSP 串流畫面，如圖 5 所示，計算單元就可持續性獲得影像訊號作為輸入資料運行計算單元上的主程式，運算結果輸出顯示在 MFD 畫面上；不斷電系統(Uninterruptible Power Supply, 簡稱 UPS)為船上所有重要設備提供緊急電力，若停

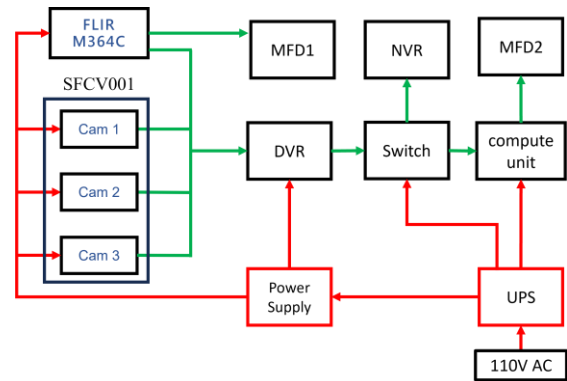


圖 1、系統架構圖。



圖 2、SFCV001 輔助避碰警示功能 Cam。



圖 3、FLIR M364C 熱成像儀。



圖 4、BOXER-6840-CFL。

電/電壓不穩時則支援電力至設備自動關機為止，防止突發斷線損耗設備。

此外，獨立於本系統，船上經常性配置之 NVR 係負責儲存船內所有監控 Cam 之影像，以及備份計算單元輸出之運算影像結果。

### 2. mAP (mean Average Precision)

為評斷物件偵測器的效能指標，本文採學術界指標 mAP (mean Average Precision) 做為依據，其利用 *IoU* (Intersection over Union) 概念評估物件偵測模型的訓練結果與測試集的 GT (Ground Truth) 偵測結果是否相近，且利用混淆矩陣中的 Precision 和 Recall，描繪出相應的 P-R 曲線 (Precision-Recall curves)，計算模型在不同的類別下的 AP (Average Precision) 值，再經平均計算 mAP 數值，當 mAP 數值越高代表模型偵測器能力表現越佳。

*IoU* 的計算方法如式(1)所示，物件偵測結果和 GT 之偵測框重疊面積越大，則 *IoU* 的數值越高。

$$IoU = \frac{\text{偵測框交集面積}}{\text{偵測框聯集總面積}} \quad (1)$$

吾人令 *IoU* 的閾值設定為 0.5，故  $IoU \geq 0.5$  則認定偵測器正確偵測到物件，依據此概念計算辨識結果的混淆矩陣 (Confusion Matrix)，如表 1 所示，並透過式(2)與式(3)計算 Precision 和 Recall 的值，逐一計算不同偵測框數量的 Precision 和 Recall，描繪出相應 P-R 曲線，如圖 6 所示。

AP 即是 P-R 曲線下的面積占比，而 AP 只針對單一類別計算，對不同的類別計算後，再進行平均，即獲得 mAP。

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

### 3. 標籤影像資料集

為使物件偵測模型具備優秀的辨識準確率和泛化能力，標籤影像資料集的準備十分重要，熱成像 raw data 的蒐集取自高雄港、基隆港，然後利用輔助標籤工具，自動處理大量 raw data 為預標籤資料，再經人工資料清洗 (data cleaning) 獲得可訓練標籤資料，最終標籤資料分為三組：training, validation, and test datasets，其中 test dataset 又做為 GT (Ground Truth) datasets，特別基於幾點挑選：(1) 應用場景泛用性、(2) 不參與訓練、(3) 具有一定物件數量、(4) 物件具關鍵代表性。本文物件類別為船隻 (boat) 浮標 (buoy) 共二種，標籤影像資料集總共累積 14,282 筆。

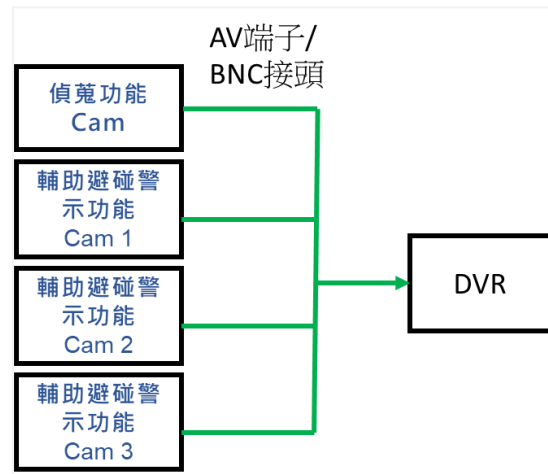


圖 1、鏡頭與 DVR 連接方式示意圖。

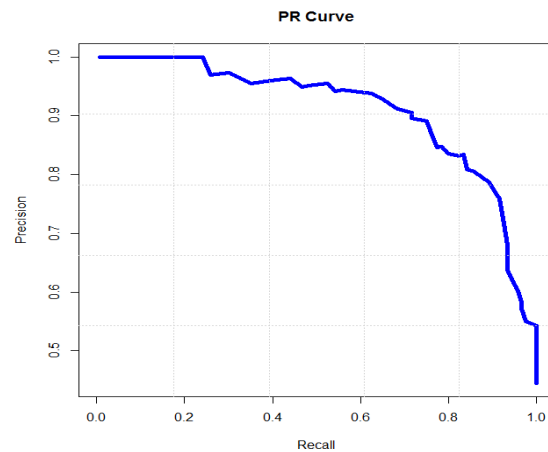


圖 6、P-R 曲線。

表 1、混淆矩陣 (Confusion Matrix)。

	實際 YES	實際 NO
預測 YES	TP (True Positive)	FP (False Positive)
預測 NO	FN (False Negative)	TN (True Negative)

### 4. 熱成像物件辨識演算法

為量化熱成像物件辨識演算法效能，比較加入新標籤資料的物件偵測器與舊有物件偵測器的辨識能力，藉結合高雄港、基隆港訓練新的物件偵測器模型；針對物件偵測器對 test dataset 數值量化辨識準確率指標採用 mAP (mean Average Precision)，在訓練模型階段，以 6:2:2 比例切分資料集為 training,

validation, and 未經訓練之 test dataset(亦稱為 GT dataset)；物件偵測器採用兼具穩定性與執行效率之 YOLOv5，模型 No.1 為基隆港周邊的三座漁港混合訓練的模型，而模型 No.2 是基於模型 No.1 擴充高雄港資料，上述兩模型經訓練後均給 test dataset 計算 mAP：模型 No.1 的 mAP 為 90.05%，模型 No.2 的 mAP 為 92.10%，整體效能提高 2.05%；圖 7~9 是模型 No.2 辨識 test dataset 之結果，效能提升的主因除了增加標籤資料外還經過完全人工重新整理，剔除 AI 訓練中易讓模型疑惑的資料，可有效地提升偵測器的準確率，且大幅減少誤判的機率。

## 5. 幾何測距演算法

本文利用幾何原理作為測距方法，首先須定義相機座標如圖 10(a)與世界座標如圖 10(b)。由於世界座標為三維座標，相機座標為二維座標且不具備深度資訊，因而無法推算存在二維座標裡物體在世界座標的位置，因此，以下將分為側視圖(圖 11)與俯視圖(圖 12)並逐步說明計算過程。

首先須假設物體位在一平坦的平面上，像是地面或是海上，我們可繪製單目鏡頭成像圖，在圖 11 及圖 12 中， $f$  為鏡頭焦距， $\alpha$  為視野範圍，這兩個參數皆可從鏡頭的規格書或對鏡頭標定得到，鏡頭的俯仰角 $\theta_p$ 及高度  $H$  可在設置鏡頭時自由調整而獲得。

當焦距或視野範圍只有其中之一時，在圖 11 中，我們有這兩者之間的幾何關係，所以水平焦距為

$$f_y = \frac{h}{2 \tan \frac{\alpha_v}{2}} \quad (4)$$

而 $\alpha_v$ 水平視野範圍為

$$\alpha_v = 2 \cdot \tan^{-1} \frac{h}{2f_y} \quad (5)$$

同理，在圖 12 中，我們也能得到鉛直焦距為

$$f_x = \frac{w}{2 \tan \frac{\alpha_h}{2}} \quad (6)$$

而 $\alpha_h$ 鉛直視野範圍為

$$\alpha_h = 2 \cdot \tan^{-1} \frac{w}{2f_x} \quad (7)$$

假設偵測到一個物體，其距離鏡頭為  $D_y$ ，透過鏡頭成像後，其成像位置的鉛直座標為  $p_y$ ，則可得到

$$D_y = H \cdot \tan(\theta_p - \beta) \quad (8)$$

而  $\beta$  可在成像平面中找到一同位角，由於同位角相等，若影像高度為  $h$ ，則

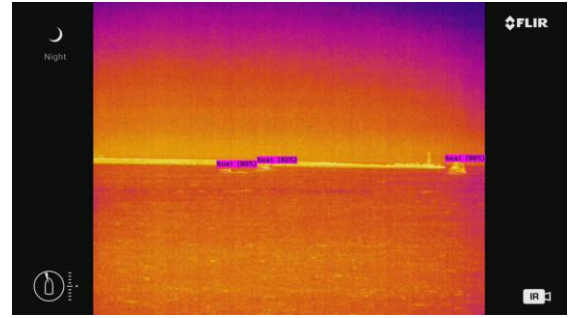


圖 7、模型辨識 test dataset 影像成果(1)。

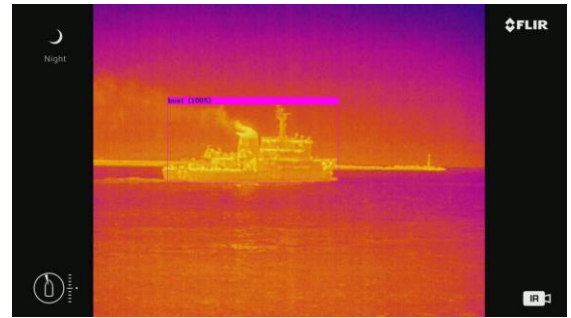


圖 8、模型辨識 test dataset 影像成果(2)。



圖 9、模型辨識 test dataset 影像成果(3)。



圖 10、(a)影像中的相機座標；(b) 物件平面世界座標之俯視圖。

$$\beta = \tan^{-1} \frac{\frac{h}{2} - p_y}{f_y} \quad (9)$$

因此我們可以將 Y 軸的距離寫作

$$D_y = H \cdot \tan(\theta_p - \tan^{-1} \frac{\frac{h}{2} - p_y}{f_y}) \quad (10)$$

在圖 11 中， $\alpha_v$  為水平視野範圍，透過鏡頭成像原理，其成像位置的水平座標  $p_x$ ， $\theta_y$  為物體的方位角，而由內錯角相等的性質，可得到

$$D_x = D_y \cdot \tan \theta_y \quad (11)$$

而  $\theta_y$  一樣可在在成像平面中找到一同位角，由於同位角相等，若影像寬度為  $w$ ，則

$$\theta_y = \tan^{-1} \frac{\frac{w}{2} - p_x}{f_x} \quad (12)$$

因此我們可以將 X 軸的距離寫作

$$D_x = H \cdot \tan(\theta_p - \tan^{-1} \frac{\frac{h}{2} - p_y}{f_y}) \cdot \tan^{-1} \frac{\frac{w}{2} - p_x}{f_x} \quad (13)$$

最終我們可得到目標物的平面世界座標

$$(x, y) = (D_x, D_y) \quad (14)$$

而目標物的距離即為

$$D = \sqrt{D_x^2 + D_y^2} \quad (15)$$

「幾何測距演算法」，熱成像測距結果如圖 13 所示，畫面右下浮標對應圖 14。估測距離 75.6 公尺，實際距離約 74.77 公尺，誤差約 1.11%。影像測距誤差來自「海水潮差」高低導致鏡頭與海平面之距離  $H$  數值變化，另一方面，水上「波浪」起伏亦導致物件與鏡頭未必精準地處於一固定平面上，另外物件辨識的「偵測框」存在跳動導致物件座標失準也是產生誤差的一大原因，以上問題將於後續章節「抗運動干擾」探討解決。總結本節結果，足見幾何測距演算法之可行性，熱成像最遠可測達約 800 公尺。

## 6. 非穩態平台之影像校準

船舶相關水上載具於海上航行時會因波浪等因素造成船體晃動，連帶影響相機拍攝之影像視角，本文採 9 軸陀螺儀協助畫面校準。陀螺儀裝置於熱成像設備上，即時回饋相機的姿態角，分別是俯仰角(Pitch)、偏航角(Yaw)以及滾動角(Roll)，其中偏航角(Yaw)僅影像畫面的水平方向左右偏移不影響測距演算法的計算，故針對俯仰角(Pitch)和滾動角(Roll)對畫面進行影像水平校準處理，目標使海平線/水平線能時刻保持在畫面的正中心或是特定水平高度，以配合演算法進行距離估算。

## 7. 抗運動干擾之水上物件定位影像測距系統

本文結合熱成像物件辨識演算法、非穩態平台影像校準以及測距演算法，開發一水上物件定位影像測距系統，但如前述提及像是天氣條件、水面的反射、波浪的遮蔽、水流造成的船舶運動和姿態變

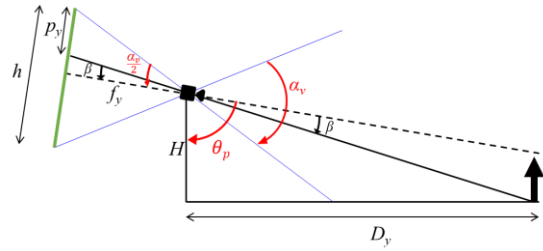


圖 11、物件成像之側視圖。

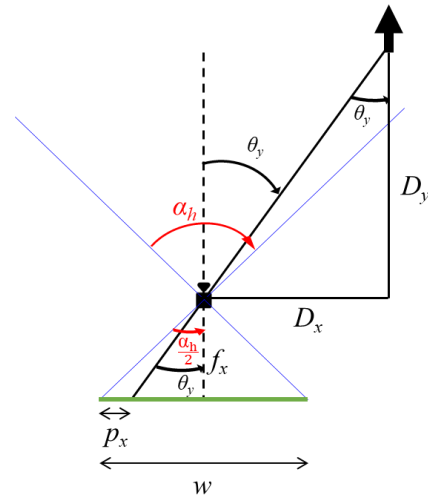


圖 12、物件成像之俯視圖。



圖 13、幾何成像演算法之估測結果。

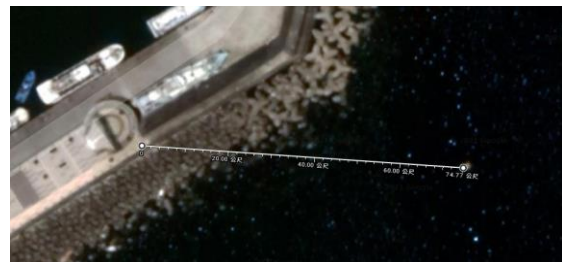


圖 14、Google 衛星地圖實測浮標距離。

表 2、對水上物件測距結果執行平均濾波。

影格組	第1組					第2組				
影格	1	2	3	4	5	6	7	8	9	10
時間(秒)	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
距離(公尺)	300	301	302	301.5	298	302.7	303.5	304	305.6	306.6
最終距離(公尺)	<b>300.5</b>					<b>304.48</b>				

化過於劇烈等環境因素干擾所帶來的判讀誤差等問題，故再增加一抗運動干擾演算法，使之具抗運動干擾之水上物件定位影像測距系統，其完整運作方法係：步驟(A)使計算單元依據熱成像鏡頭拍攝的影像進行物件偵測 (Object detection)，並確認水上物件位於該影像中的位置；接著步驟(B)由計算單元算出步驟(A)中的複數幀影格裡水上物件分別的距離，並將該複數幀影格依照一時間區段劃分為複數幀影格組；最後，步驟(C)中，計算單元將每個複數幀影格組中的每個複數幀影格所綜合計算出一水上物件的距離並進行抗運動干擾之濾波，以獲得水上物件的最終距離。

舉例而言，在熱成像鏡頭的影格速率假設固定為 10 FPS 且每 5 幀影格劃分為一個影格組的前提下，會得到兩個影格組，計算單元每秒會需要計算 10 次水上物件的距離並且針對兩個影格組分別平均濾波 (Mean Filter) 以獲得水上物件的分別兩個影格組最終距離之測定，範例計算如表 2，即在 1 秒內會對水上物件顯示兩次的最終距離 (前 0.5 秒顯示為 300.5 公尺；後 0.5 秒顯示為 304.48 公尺)。每個影格組的最終距離係該影格組中每幀影格所計算出距離的平均值。有關於濾波(filter)方法的部份可獨立替換其他方法以獲得最合理的抗運動干擾測距結果，目前系統測距能力能維持在誤差±10%以內；未來將改使用卡爾曼濾波器(Kalman filter)以更貼近真實測距數值。

## 8. 船舶避碰警示系統

延續前小節「抗運動干擾之水上物件定位影像測距系統」，該系統進一步根據演算法所計算出之物件距離數值大小分為遠、中、近三種等級，其警示規則如表 3 所示，輔助使用者根據不同的狀況做出決策。圖 15、16 別為高雄港船隻上所拍攝畫面之三級距離呈現。

## 9. 使用者 UI 介面開發

本文利用 Python 開發使用者介面，如圖 17 所示，整合 DVR 取得四組熱成像鏡頭並結合物件偵測

表 3、避碰警示判定規則表。

距離代號	顏色	距離範圍
近(Near)	紅	<200m
中(Mid)	綠	200m~900m
遠(Far)	藍	>900m

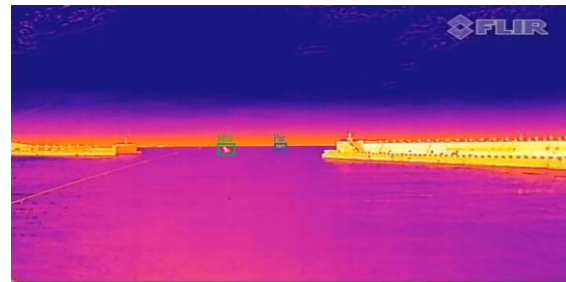


圖 15、船上所拍攝畫面(中、遠)。

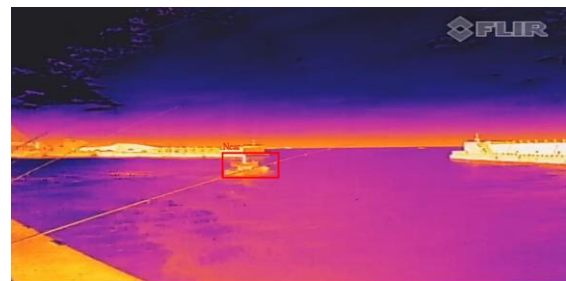


圖 16、船上所拍攝畫面(近)。

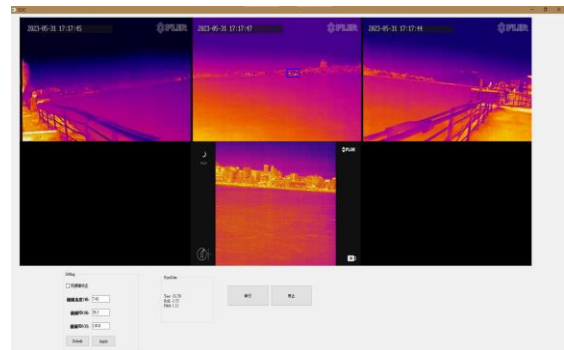


圖 17、使用者介面示意圖。

器、測距演算法、非穩態平台之影像校準、抗運動干擾濾波、避碰警示提示音、介面設計為一般船避碰警示系統，並稱為**科力航太智慧夜視整合系統**，該系統已實裝在航港局預計 113 年 4 月交船之「航港 1 號」，期望落實智慧航安監控船舶的實地場域應用。

## 10. 可見光攝影機的測距系統

順應國際著重自駕船的趨勢，SOIC 亦開發一套輔助自動駕駛系統裝置於遊艇上，於 2022 年試航，本文之**科力航太智慧夜視整合系統**可見光鏡頭版也安裝該船上測試，該系統的核心算法相同，無論熱成像或可見光鏡頭均可適用，實測結果如圖 12~14，當系統感知到存在物件進入(far, middle, near)三種測距範圍中，即可回饋相關資訊給自動駕駛系統，以演算決策迴避碰撞策略，這將可大幅幫助自動駕駛系統於電腦視覺感測的能力。

## 結論

本文重要成果分三部分 1. **熱成像物件辨識演算法**：使用影像資料集總張數為 14,282 張，最終成果模型之 mAP 為 94%，為計算由以下二項類別 AP 之平均值所得出(boat (88.35%)及 buoy(99.92%))；2. **抗運動干擾之水上物件定位影像測距暨避碰警示系統**：利用相機內外參數：鏡頭高度、鏡頭俯仰角、焦距及 FOV 等，將相機座標系轉換為世界坐標系，緊接進行非穩態平台之影像校準與抗運動干擾之濾波獲得最終測距結果；3. **科力航太智慧夜視整合系統**：將上述成果整合進一使用者介面，供使用者查看系統運作結果，包含攝影機畫面、物件辨識結果、與物件之距離(遠、中、近)等。

## 未來展望

許多的海上不法事件為避免行動曝光，通常關閉船隻 AIS 及 VMS 等發送座標信號系統，故增進執法單位主動式偵蒐能力十分關鍵。有鑑於國際許多組織已籌備彙整 AIS、VMS、雷情等船端偵蒐資訊，於岸端圖台上疊圖顯示以利勤務指揮已普及。本文**科力航太智慧夜視整合系統**具備的物件追蹤功能，同個水上物件之動向可以同步傳輸至岸端更新至圖台上即時顯示，對於機關勤務之偵蒐將降低人工紀錄與手動傳輸的程度，自動化的追蹤功能降低艦船人員負擔並得以兼顧海上其他臨時突發狀況，岸端勤指中心亦能同步船端現況以精進調度評估。

## 參考文獻

1. "Introducing the machine odyssey, the world's first 1,000 nm autonomous voyage," *sea machines*, 2021.



圖 12、自駕船上所拍攝畫面(Mid, Far)。



圖 13、自駕船上所拍攝畫面(Mid)。



圖 14、自駕船上所拍攝畫面(Near, Far)。

2. "Machine Vision to assist competitors' collision avoidance in the Transat Jacques Vabre," *SEA.AI*, 2023.
3. "UC-win/Road Ship Handling Simulator," *FORUM8*.
4. "Finferries' Falco world's first fully autonomous ferry," *Finferries*, 2018.
5. "A significant outcome of the collaboration between four maritime stakeholders from different sectors demonstrates the feasibility of an AI-based autonomous navigation system," *Korean Register*, 2023.
6. S. Youm, C. Kim, S. Choi, and Y. S. Kang, "Development of a methodology to predict and monitor emergency situations of the elderly based on object detection," *Multimed. Tools Appl.*, vol. 78, no. 5, pp. 5427–5444, 2019.
7. W. Zhu, W. Huang, Z. Lin, Y. Yang, S. Huang, and J. Zhou, "Data and feature mixed ensemble based extreme learning machine for medical object detection and segmentation," *Multimed. Tools Appl.*, vol. 75, no. 5, pp. 2815–2837, 2016.
8. Z. Ren, X. Zhang, and S. Yang, "A Real-Time Suspicious Stay Detection System Based on Face Detection and Tracking in Monitor Videos," in *2017 10th International Symposium on*

- Computational Intelligence and Design (ISCID)*, 2017, vol. 1, pp. 264–267.
9. R. Girshick, “Fast r-cnn,” in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.
  10. 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*, 2015, pp. 91–99.
  11. K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask r-cnn,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.
  12. J. Redmon and A. Farhadi, “Yolov3: An incremental improvement,” *arXiv Prepr. arXiv1804.02767*, 2018.
  13. A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” *arXiv Prepr. arXiv2004.10934*, 2020.
  14. Github: Yolov5. <https://github.com/ultralytics/yolov5>.
  15. W. Liu et al., “Ssd: Single shot multibox detector,” in *European conference on computer vision*, 2016, pp. 21–37.
  16. M. Tan, R. Pang, and Q. VLe, “Efficientdet: Scalable and efficient object detection,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10781–10790.
  17. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.