

國立臺灣師範大學科技與工程學院電機工程學系

碩士論文

Department of Electrical Engineering

College of Technology and Engineering

National Taiwan Normal University

Master Thesis

結合頭部姿態估計與補償的視線追蹤

Gaze Tracking with Head Pose Estimation and
Compensation



陳璽文

Xi-Wen Chen

指導教授: 高文忠 博士

Advisor: Wen-Chung Kao Ph.D.

共同指導教授: 陳建隆 博士

Co-Advisor: Jann-Long Chern Ph.D.

中華民國 114 年 1 月

January, 2025

致謝

時光飛逝，碩士這段旅程接近尾聲，首先要感謝我的指導老師。感謝電機系高文忠老師，給了我極大的自由去探索研究，讓我可以盡情發揮自己的創意；同時也感謝數學系陳建隆老師，支持我跨領域進入電機研究所，並在這段路上給予我重要的建議和信任。您們的鼓勵和指導，是我完成碩士論文的重要支柱。

接著，我要感謝一路陪伴我的夥伴與朋友們。中央的育愷學長，從大學到碩士期間，您給了我許多方向上的指引，讓我少走了很多彎路；在 SocLab 的實驗室中，忠謀、少榆、翔元、秋貴、陳灝、宇彤和鈺珊，感謝你們在這兩年裡與我共同努力，熬夜實驗、討論數據，也分享了無數的歡樂與笑聲，這些回憶將永遠珍藏在我心中。

除此之外，我也要感謝我的大學好友建宇，從求學路上到現在，你一直是我最大的後盾，每當遇到困難時總能帶給我許多幫助。還有我的女朋友昀亭，謝謝你在我忙碌又充滿壓力的碩士生活中給予理解與支持，默默陪伴我度過最困難的時刻，讓我能夠專注於研究工作。

最後，我要感謝我的家人，永遠支持我的選擇，包容我的一切。這份愛與信任，讓我能無後顧之憂地專注完成碩士學業。我的研究專注於可見光眼動儀，提升其在頭部轉動時的精確度與穩定度。這一路充滿挑戰，但也因為有這麼多人陪伴與鼓勵，我才能不斷突破自我，順利完成這份屬於我的學術旅程。

結合頭部姿態估計與補償的視線追蹤

學生：陳璽文

指導教授：高文忠 教授
陳建隆 教授

國立臺灣師範大學電機工程學系碩士班

摘 要

本文提出了一種基於可見光影像的視線追蹤系統，採用單一高速相機，取代傳統依賴紅外光源或專用傳感器的方案，從而顯著提升了使用者體驗。然而，這種設置在補償頭部移動方面面臨更大的挑戰。為解決此問題，我們設計了一種新型視線追蹤系統，結合了精確的頭部姿態估計方法。該方法通過識別臉部特徵點並解決 2D 到 3D 的對應問題，獲取特徵點的 3D 坐標，進而估算頭部運動。該系統能夠實時更新眼球模型並準確計算虹膜區域的初始位置。實驗結果表明，當使用者進行輕微頭部移動或旋轉時，該系統能有效提高視線追蹤的精度與準確性。

關鍵字：凝視追蹤、頭部姿態估計、3D 眼球模型、深度學習

Gaze Tracking with Head Pose Estimation and Compensation

student : Xi-Wen Chen

Advisors : Dr.Wen-Chung Kao
Dr.Jann-Long Chern

Department of Electrical Engineering
National Taiwan Normal University

ABSTRACT

This paper proposes a visible-light-based gaze tracking system that utilizes a single high-speed camera, replacing traditional systems that rely on infrared light sources or dedicated sensors, thereby significantly enhancing user experience. However, this configuration poses greater challenges in compensating for head movements. To address this issue, we designed a novel gaze tracking system that integrates an accurate head pose estimation method. The method identifies facial feature points and resolves the 2D-to-3D correspondence problem to obtain the 3D coordinates of these points, which are then used to estimate head motion. The system is capable of real-time updates to the eye model and precise calculation of the initial position of the iris region. Experimental results demonstrate that the system effectively improves gaze tracking accuracy and precision when users perform slight head movements or rotations.

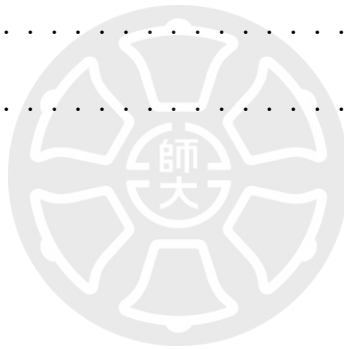
Keywords: Gaze Tracking, Pose Estimation, 3D Eye Model, Deep learning

目錄

	Page
致謝	i
Abstract	ii
目錄	iv
圖目錄	vii
表目錄	ix
第一章 緒論	1
1.1 研究背景	1
1.2 研究問題及目的	2
1.3 研究目標	3
第二章 文獻探討	4
2.1 3D 眼球模型與虹膜匹配	4
2.2 眼球中心定位	9
2.3 頭部姿態估計	12
2.3.1 基於幾何特徵的方法	12
2.3.2 基於深度學習的方法	13
2.3.3 基於三維模型的方法	14
2.4 虹膜分割模型	15
2.5 螢幕與視線的映射公式	17
2.5.1 映射關係建立方式	18
2.5.2 回歸分析	19
2.5.3 偵錯點方法	20

第三章 系統架構與設計	22
3.1 臉部特徵點的定位與提取	22
3.2 基於頭部姿態的眼球中心校正	25
3.2.1 頭部姿態估算技術	25
3.2.2 眼球中心初始計算	28
3.2.3 動態校正方法	29
3.3 虹膜區域分割	30
3.3.1 虹膜區域的偵測方法	31
3.3.2 分割後的數據清理	33
3.3.3 虹膜參數提取	34
3.4 虹膜特徵匹配	35
3.4.1 搜尋範圍優化	36
3.4.2 算分區域	38
3.4.3 虹膜匹配	41
3.5 視線方向與螢幕位置的映射關係	42
第四章 實驗結果	45
4.1 實驗環境與系統設置	45
4.2 頭部姿態估計實驗	46
4.3 虹膜分割模型實驗與結果	48
4.3.1 數據集處理與修改	48
4.3.2 模型預處理和訓練	49
4.3.3 量化評估與結果	49
4.4 虹膜匹配優化實驗	51
4.4.1 階層搜尋方法	51

4.4.2 搜尋範圍優化對計算效率的影響	51
4.5 凝視估計系統性能評估	52
4.5.1 像素視角轉換	52
4.5.2 系統精準度	52
4.5.3 系統精密度	53
4.6 熱區圖結果對比	53
4.6.1 九點驗證	53
4.6.2 十七點驗證	57
4.6.3 映射方法在校準的差異	60
第五章 結論與未來展望	63
5.1 結論	63
5.2 未來展望	63
References	65
自傳	68
學術成就	70



圖目錄

1.1 頭部旋轉不同角度時眼睛的變化。[1]	2
2.1 3D 眼球模型	5
2.2 直視相機時，虹膜邊界點 $P_I^0(\alpha)$ 與 α 的關係示意圖	6
2.3 左圖為虹膜中心與邊界點之間的幾何關係，右圖為邊界點上的步進是意圖。	8
2.4 更精準的眼角影像	10
2.5 內眼角定位眼球模型的示意圖，展示了內眼角與眼球幾何中心的相對位置及模型匹配過程。	11
2.6 頭部在三維空間中的旋轉方向 [2]	12
2.7 基於 ResNet 的頭部姿態估計架構 [3]	13
2.8 重建 3D 頭部形狀和紋理的處理步驟 [4]	15
2.9 RITnet[5] 的架構細節與數據集影像差異示例。	16
2.10 DCNN 網路架構 [6]	17
2.11 使用者觀看螢幕九點示意圖，	18
2.12 不同使用者觀看螢幕九點視線分布，可以觀察到會有錯誤點的產生。	21
3.1 系統架構圖	22
3.2 mediapipe 臉部特徵點	24
3.3 篩選出來的 14 個臉部特徵點。	26
3.4 PnP 示意圖：3D 點透過相機姿態 (R, T) 投影到 2D 成像平面。	27
3.5 眼球位置更新的流程：通過旋轉 R 和平移 T 從頭部模型空間轉到相機空間，再經由 K 映射至圖像空間，最終確定眼球中心 P_{eye} 的位置。	30
3.6 左圖為輸入的眼睛影像，右圖為模型輸出的虹膜分割結果。	33
3.7 虹膜分割後處理：左圖為處理前，具有一些雜訊，虹膜輪廓較不清晰；右圖為處理後，濾掉雜訊且虹膜輪廓較平滑。	34
3.8 後處理步驟提取虹膜初始中心與半徑。	35
3.9 虹膜搜尋範圍示意圖。搜尋範圍越大或搜尋精度縮小，都會導致運算瓶頸。	37
3.10 眼睛往左邊看虹膜的與影像上的情況。	38
3.11 使用者直視相機時所定義的算分區域	39

3.12 邊界算分方式，根據水平偏移量 R_x 以及垂直偏移量 R_y 與虹膜初估半徑的對數值之間的比值來決定。	40
3.13 根據虹膜中心與眼球中心的相段關係所設置的補償算分方式，決定當前可信任的邊界。	41
3.14 映射公式判斷流程圖	44
4.1 XIMEA MQ042MG-CM	45
4.2 環境配置圖	46
4.3 數據集 ground truth 修改示例。上排為原始標註，僅包含虹膜；下排為修改後標註，將瞳孔與虹膜視為一致。	48
4.4 測試資料虹膜分割結果圖	50
4.5 驗證階段觀看九點的使用者狀態，	54
4.6 data2：前後的熱區圖結果對比。	55
4.7 data3：前後的熱區圖結果對比。	55
4.8 data6：前後的熱區圖結果對比。	56
4.9 data7：前後的熱區圖結果對比。	56
4.10。	57
4.11 data19：熱區圖比較	58
4.12 data28：熱區圖比較	58
4.13 data24：熱區圖比較	59
4.14 data29：熱區圖比較	60
4.15 data5：校準階段不同映射方式的差異。	61
4.16 data5：校準階段不同映射方式的差異。	61

表目錄

1.1 三種視追蹤方法的優勢與挑戰	3
3.1 旋轉角度搜索範圍比較	38
4.1 四種數值演算法在 solvePnP 問題中的性能比較	47
4.2 改進的 RITnet 模型在各數據集上的表現	50
4.3 階層搜尋方法參數設置	51
4.4 搜尋範圍優化對計算時間的影響	52
4.5 Data 2 精確度與精準度比較	55
4.6 Data 3 精確度與精準度比較	55
4.7 Data 6 精確度與精準度比較	56
4.8 Data 7 精確度與精準度比較	57
4.9 Data19 驗證階段精確度與精準度比較	58
4.10 Data28 驗證階段精確度與精準度比較	58
4.11 Data24 驗證階段精確度與精準度比較	60
4.12 Data29 驗證階段精確度與精準度比較	60
4.13 Data5 驗證階段精確度與精準度比較	61
4.14 Data 10 驗證階段精確度與精準度比較	62

第一章 緒論

1.1 研究背景

隨著人機互動技術的迅速發展，視線追蹤技術在數位實境（Digital Reality）、人體工學以及生物識別等領域中的應用愈發普及。視線追蹤技術能夠精確地偵測並解析使用者的視線方向，為人機互動提供更加自然、直觀的操作方式，並在多個領域中展現出其重要價值。例如，在增強實境（AR）和虛擬實境（VR）中，視線追蹤技術可提升交互體驗，使系統更具沉浸感；在醫療領域中，它可以用於患者注意力診斷、康復訓練與手術輔助；而在教育領域，它能實現對學習者注意力的監控與評估，促進個性化教學的實現。

然而，目前主流的視線追蹤系統大多依賴紅外線（IR）光源技術來獲取穩定的瞳孔追蹤效果。這種技術在短時間內確實能有效提升追蹤的準確性和穩定性，但同時也伴隨一些顯著的限制。一方面，紅外線光源的長時間使用可能導致使用者的不適，特別是在需要長時間應用的場合中，例如日常辦公或娛樂；另一方面，紅外線視線追蹤系統通常需要配備專門的硬體設備，這大幅增加了系統的成本，使其在普及化應用中面臨挑戰。此外，對於便攜式與輕量化的應用場景，例如穿戴式設備，紅外線光源的硬體需求也增加了設計與開發的複雜性。

為了解決上述問題，基於可見光（Visible Spectrum）的視線追蹤技術應運而生。該技術通過利用自然光和普通的可見光相機，避免了對專用紅外光源的依賴，從而有效降低了硬體成本，並改善了使用者的舒適度。此外，可見光視線追蹤技術還具有較強的適應性，可直接應用於現有的多數消費級攝影設備上，例如手機、筆記型電腦和桌上型電腦的內建攝像頭，為長期應用和日常使用提供了更為經濟高效的解決方案。

隨著人們對互動設備使用舒適度、便攜性以及成本效益的要求日益提升，基於可見光的視線追蹤技術逐漸成為視線追蹤領域的研究熱點。其不僅能突破傳統

紅外線技術的限制，還有望在未來成為下一代視線追蹤技術的核心基礎，推動人機互動技術在更多領域實現創新應用。

1.2 研究問題及目的

然而，在可見光條件下進行精確的視線追蹤具有更高的挑戰性。首先，可見光系統無法單純依賴瞳孔的高對比特徵進行追蹤，而必須精確辨識虹膜邊界，才能推算出眼睛的注視方向。其次，日常使用中使用者的頭部動作幾乎無法避免，而頭部的移動或轉動往往會顯著影響眼動儀的精確度與穩定性。如 Figure 1.1 所示，雖然在三種不同的頭部轉動角度下，視線方向保持一致，但眼睛外觀卻隨角度變化而發生顯著差異，增加了系統辨識的難度。

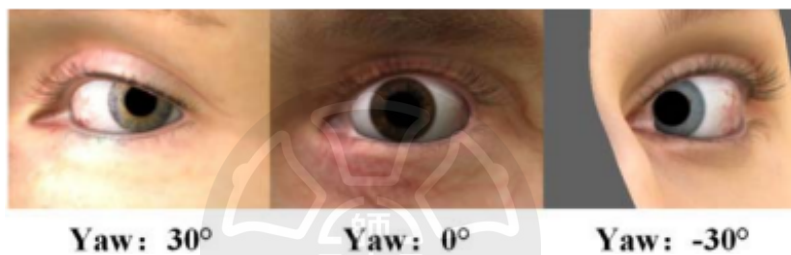


Figure 1.1: 頭部旋轉不同角度時眼睛的變化。[1]

現有的可見光視線追蹤技術根據視線方向估計的方法主要分為基於“2D 特徵”、“3D 眼球建模”和“深度學習數據處理”三大類，如表 1.1。在 2D-Based 方法中，系統通過捕捉眼部圖像中的 2D 特徵（如瞳孔中心、虹膜邊緣或角膜反射點），利用幾何關係估算視線方向。該方法的硬體需求較低，實現簡單，在靜態環境中表現出色，但對頭部運動和光照變化的敏感性限制了其在動態場景中的應用。相比之下，3D-Model-Based 方法通過構建眼球的幾何模型，利用三維空間中的眼球姿態與視線方向的物理關係進行估算，理論上能更好地補償頭部運動對視線估計的影響。然而，這類方法通常需要多攝像機系統或高精度的光學元件，對硬體配置要求較高，且在光照條件多變或頭部旋轉角度較大時仍可能存在性能下降的問題。近期，深度學習數據處理方法在視線追蹤領域得到了廣泛應用，特別是在可見光環境下的挑戰中展示出優勢。這類方法通過大規模數據集進行模型訓練，自動學習眼部特徵與視線方向的非線性關係，無需明確構建幾何模型或依

賴多攝像機系統，從而提高了對複雜光照條件和頭部運動的適應能力。然而，深度學習方法對標註數據集的依賴性較強，訓練過程需要高計算資源支持，並且在某些情況下對視線方向的估計可能缺乏幾何方法的可解釋性。

基於以上對現有視線追蹤技術方法的綜述，可以明確看出各類方法的優劣勢以及現存的挑戰。傳統的 2D-Based 方法因其簡單性而適合靜態場景，但在動態場景中穩定性不足；3D-Model-Based 方法雖然精度高，但硬體需求和環境適應性存在局限；深度學習方法則在光照和頭部運動的變化中表現出色，但受制於數據和資源的需求。

Table 1.1: 三種視追蹤方法的優勢與挑戰

方法	優勢	挑戰
2D-Based	硬體簡單，靜態環境性能優異	不適應頭部運動，對光照敏感
3D-Model-Based	補償頭部運動精度高	硬體需求高，光照條件敏感
深度學習方法	適應光照與頭部運動，無需多攝像機	訓練成本高，依賴大數據

1.3 研究目標

為了解決上述研究問題，本研究旨在提出一種融合 2D 和 3D 方法優勢的視線追蹤技術框架，並探討深度學習技術在可見光環境中的應用潛力。具體而言，本研究的目標包括以下三點：首先，構建一種多層次的視線估算模型，結合 2D 與 3D 方法的優勢，在硬體成本與性能之間達成平衡，並提升系統在動態場景中的穩定性；其次，開發深度學習輔助的數據處理與建模方法，實現系統對光照條件變化的強適應能力；最後，通過引入幾何建模技術，著重構建眼球與視線方向之間的數學關係，增強模型對視線估算過程的解釋性。

第二章 文獻探討

在視線追蹤技術領域，現有研究與應用多集中於 2D 與 3D 方法的開發與比較。然而，隨著應用場景日益多樣化，如何有效結合這兩種方法的優勢，並在硬體成本與性能之間取得平衡，成為一個亟待解決的挑戰。此外，視線追蹤技術在可見光環境中的穩定性和適應性仍然存在諸多限制，特別是在動態場景和光照條件變化的情境下。深度學習技術的發展為視線追蹤領域提供了新的機遇，其在數據處理與建模中的應用潛力，尤其是對複雜環境的適應能力，吸引了越來越多的關注。

本章將回顧與本研究相關的文獻，重點聚焦於以下幾個方面：3D 眼球模型與虹膜區域的匹配技術，包括眼球中心定位與建模公式的推導（詳見第 2.1, 2.2 章）；頭部姿態估計技術，涵蓋基於幾何特徵、深度學習及三維建模的方法（參見第 2.3 章）；以及第 2.4 章，關於虹膜分割模型的應用。同時，還將探討螢幕與視線的映射公式，包括映射關係的建立方式、回歸分析方法以及傾錯點策略（參考第 2.5 章）。通過對這些領域文獻的梳理與分析，為本研究的理論基礎與方法框架提供支持，並進一步明確現有研究的局限性與改進方向。

2.1 3D 眼球模型與虹膜匹配

在估算人眼視線方向的研究中，眼球模型的選擇至關重要。過去的研究中，存在多種眼球模型的構建方式，例如基於真實解剖結構的複雜模型，或是將眼球分解為多層結構的分層模型。然而，這些方法通常因模型的高複雜度和計算需求，難以在實際應用中達到平衡。

為了應對這些挑戰，許多研究採用了簡化眼球模型的策略。其中，最常見的方法是將眼球假設為一個完美的球體。這種球形模型能有效模擬眼球的旋轉和運動特性，同時顯著降低模型建構和運算的複雜度，因此成為實用且廣泛採用的解決方案。如文獻 [7] 所述，該研究使用球形模型模擬眼球的運動，如 Figure 2.1

所示，將眼球視為一個理想球體。虹膜被認為是附著於球面的一塊幾何區域。在眼睛正對攝像頭時，虹膜呈現為一個正圓形；而當眼球轉動至其他方向時，虹膜的形狀會根據觀測角度不同，表現為各種不同的橢圓形。這種模型不僅能準確描述眼球的旋轉行為，還能夠為視線追蹤系統提供高效的建模基礎。

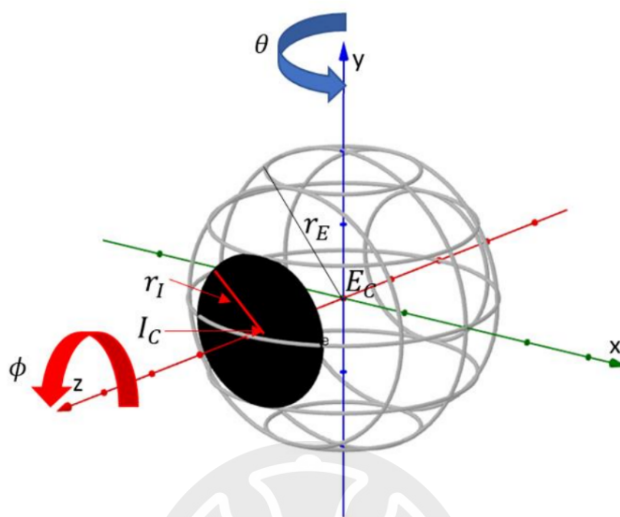


Figure 2.1: 3D 眼球模型

模型中，符號 E_c 表示眼球的中心點， r_E 為眼球的半徑； I_c 表示虹膜的中心點， r_I 為虹膜的半徑。在眼球轉動時，角度參數 θ 和 ϕ 用來描述眼球的旋轉角度。

為了進一步將 3D 眼球模型應用於影像處理，下一步需要推導眼球模型的旋轉與虹膜邊界形狀變化之間的數學關係。這一過程的核心是描述眼球在三維空間中的旋轉如何影響虹膜區域在二維影像平面上的投影。

具體而言，眼球的旋轉可由偏航角 (yaw) 和翻滾角 (roll) 來表示，也就是 Figure 2.1 的 θ 和 ϕ ，而虹膜在眼球表面的位置和形狀則可由球坐標系中的參數進行描述。當眼球轉動時，虹膜區域會從視角正對時的圓形逐漸變為橢圓形。這一形變可以通過透視投影的數學模型進行推導。

為了簡化後續的推導，我們假設攝像機鏡頭位於 z -軸上，並將眼球模型理想化為一個半徑為 1 的單位球 (即 $r_E = 1$)。在此設定下，虹膜理論上在攝像機視

角中將呈現為一個正圓形。如 Figure 2.2 所示，虹膜邊界上任意一點的位置可以用極座標表示為 $P_I^0(\alpha) = [r_I, \alpha]$ ，其中 α 表示極角， r_I 為虹膜半徑。

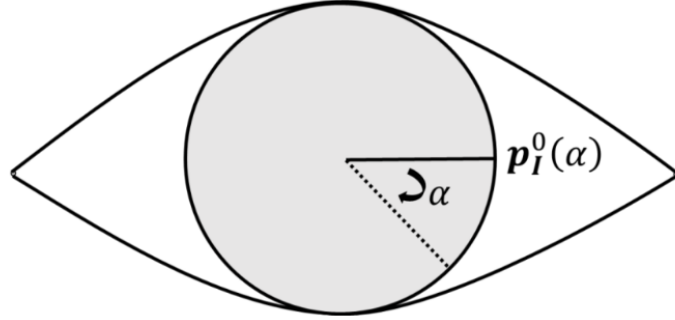


Figure 2.2: 直視相機時，虹膜邊界點 $P_I^0(\alpha)$ 與 α 的關係示意圖

單位球的數學描述可表示為：

$$x^2 + y^2 + z^2 = 1 \quad (2.1)$$

在此基礎上，假設 (x_I^0, y_I^0, z_I^0) 表示眼球正對攝像機時虹膜邊界上的任意一點，且虹膜半徑與眼球半徑的比值 $\tau = r_I/r_E$ 用於表示單位球上相對的虹膜半徑。此時， $P_I^0(\alpha)$ 的具體表示為：

$$P_I^0(\alpha) = \begin{bmatrix} x_I^0 \\ y_I^0 \\ z_I^0 \end{bmatrix} = \begin{bmatrix} \tau \cos \alpha \\ \tau \sin \alpha \\ \sqrt{1 - x_I^{0^2} - y_I^{0^2}} \end{bmatrix} = \begin{bmatrix} \tau \cos \alpha \\ \tau \sin \alpha \\ \sqrt{1 - \tau^2} \end{bmatrix}, \quad \alpha \in [-\pi, \pi].$$

當眼球發生旋轉時，虹膜邊界的位置也會隨之改變。我們可以將眼球的旋轉分解為繞 y -軸（偏航角，yaw）的旋轉角度 θ 和繞 z -軸（翻滾角，roll）的旋轉角度 ϕ 。在這種情況下，任意旋轉後的虹膜邊界可以看作是在直視情況下的虹膜邊界點經過兩個旋轉矩陣的變換。這兩個矩陣分別是繞 y -軸的旋轉矩陣 $R_y(\theta)$ 和繞 z -軸的旋轉矩陣 $R_z(\phi)$ 。

旋轉後的虹膜邊界點 $P_I(\alpha)$ 的三維座標 (x_I, y_I, z_I) 可以表達為：

$$P_I(\alpha) = R_z(\phi)R_y(\theta)P_I^0(\alpha),$$

其中， $P_I^0(\alpha)$ 表示未旋轉時的虹膜邊界點。

具體地，旋轉矩陣 $R_y(\theta)$ 和 $R_z(\phi)$ 的組合形式為：

$$R_z(\phi) = \begin{bmatrix} \cos \phi & -\sin \phi & 0 \\ \sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad R_y(\theta) = \begin{bmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{bmatrix}.$$

經過這兩個旋轉矩陣作用，最終的邊界點 $P_I(\alpha)$ 可進一步展開為：

$$P_I(\alpha) = \begin{bmatrix} (\tau \cos \theta \cos \alpha + \sqrt{1 - \tau^2} \sin \theta) \cos \phi - \tau \sin \alpha \sin \phi \\ (\tau \cos \theta \cos \alpha + \sqrt{1 - \tau^2} \sin \theta) \sin \phi + \tau \sin \alpha \cos \phi \\ -\tau \sin \theta \cos \alpha + \sqrt{1 - \tau^2} \cos \theta \end{bmatrix}, \quad \alpha \in [-\pi, \pi].$$

為了確定虹膜邊界點在影像平面上的位置，我們需要將虹膜的三維邊界點 $P_I(\alpha)$ 投影到影像平面上。這一過程是基於透視投影模型，考慮了眼球旋轉的幾何變化，並將三維空間中的點轉換為影像平面上的二維坐標。

根據透視投影原理，虹膜邊界點在影像平面上的坐標 $p_D^I(\alpha)$ 可以表示為：

$$p_D^I(\alpha) = r_E \begin{bmatrix} (\tau \cos \theta \cos \alpha + \sqrt{1 - \tau^2} \sin \theta) \cos \phi - \tau \sin \alpha \sin \phi \\ -((\tau \cos \theta \cos \alpha + \sqrt{1 - \tau^2} \sin \theta) \sin \phi + \tau \sin \alpha \cos \phi) \end{bmatrix} + \begin{bmatrix} x_C^D \\ y_C^D \end{bmatrix}.$$

在上述公式中， r_E 表示眼球半徑， τ 為虹膜半徑與眼球半徑的比值， θ 和 ϕ 分別為眼球的偏航角 (yaw) 與翻滾角 (roll)， $\alpha \in [0, 2\pi]$ 為虹膜邊界點的角度參數。 x_C^D 和 y_C^D 為影像平面上眼球中心的坐標。

此外，值得注意的是，公式中的 y 座標需要額外加入一個負號。這是由於影

像平面座標系與球面座標系的定義差異所導致的。在球面座標系中，正向的 y 軸方向與影像平面座標系的 y 軸方向相反。為了正確地將球面上的點投影到影像平面上，我們需要在轉換公式中對 y 座標加入一個負號，來校正這一方向差異，使其符合影像平面座標系的定義。

投影到影像上的虹膜邊界點可以通過影像的梯度計算 Fitness Score，用於評估候選虹膜匹配結果的準確性和穩定性。這一過程的核心在於利用影像梯度信息，評估虹膜邊界點與候選邊界點的相關性，進而計算視線估計所需的最佳參數。

首先，假設虹膜邊界點的像素坐標為 $p_0 = (L_x, L_y)$ ，並且虹膜中心位置為 (I_x, I_y) 。邊界點與中心點的角度 $\hat{\alpha}$ 可以表示為：

$$\hat{\alpha} = \tan^{-1} \left(\frac{L_y - I_y}{L_x - I_x} \right). \quad (2.1)$$

接著，通過步進方式生成候選虹膜邊界點，可以參考示意圖2.3。對於每個步進量 i ，候選點 P_i 的位置為：

$$P_i = (L_x + i \cdot \cos \hat{\alpha}, L_y + i \cdot \sin \hat{\alpha}), \quad (2.2)$$

其中， i 為步進的距離。

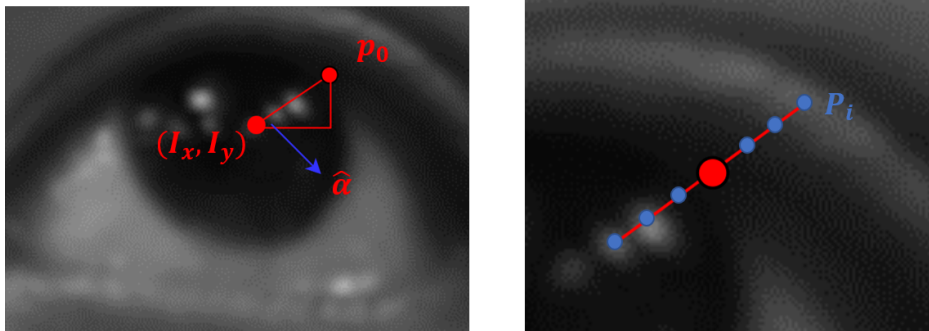


Figure 2.3: 左圖為虹膜中心與邊界點之間的幾何關係，右圖為邊界點上的步進示意圖。

在生成候選邊界點後，我們利用影像梯度計算候選邊界點與實際虹膜邊界的

相關性。梯度比值 $g(L_x, L_y, I_x, I_y)$ 的定義如下：

$$g(L_x, L_y, I_x, I_y) = \frac{\sum_{i=0}^3 \text{Img}(P_i) + 1}{\sum_{i=-3}^0 \text{Img}(P_i) + 1}, \quad (2.3)$$

其中， $\text{Img}(P_i)$ 表示影像在候選點 P_i 處的梯度值。

在獲得梯度比值後，我們可以進一步計算整體的 Fitness Score，用於評估所有候選點的匹配質量。Fitness Score 的定義如下：

$$f(\theta, \phi) = \sum_{\alpha} \log(g(L_x, L_y, I_x, I_y)), \quad (2.4)$$

其中， θ 和 ϕ 分別為眼球的偏航角和翻滾角， α 為虹膜邊界點的角度。

最後，通過最大化 Fitness Score 的方式，我們可以確定最佳的眼球旋轉參數 θ^* 和 ϕ^* ：

$$\theta^*, \phi^* = \arg \max_{\theta, \phi} f(\theta, \phi). \quad (2.5)$$

這一表達式將虹膜邊界點的旋轉與三維座標進行了精確對應，為後續的數據處理與模型應用提供了數學基礎。

2.2 眼球中心定位

準確定位眼球中心是視線追蹤系統中的關鍵步驟。眼球中心的坐標直接影響虹膜邊界的精確匹配與視線估計的整體性能。在動態場景或使用頭部運動的情況下，精確的眼球中心定位對於提升系統的穩定性和準確性尤為重要。為了實現高效且準確的眼球中心定位，研究者提出了多種基於幾何建模和深度學習的技術。

利用眼角（內眼角或外眼角）作為參考點來定位眼球中心是一種經過驗證的有效方法。眼角作為穩定的幾何特徵，能在大多數環境中保持良好的可見性和一致性，因此被廣泛應用於眼球模型的構建和匹配過程中。例如，Kao [8] 等人在

研究中提出了一種基於內眼角的三維眼球模型方法。他們利用內眼角的位置作為基準，計算眼球中心的相對位置，並通過輔助頭部補償技術，提升了系統在動態場景中的穩定性與匹配精度。該方法能有效應對頭部小幅度運動和視角偏移情境，尤其在面對不同使用者的眼部幾何特性時，表現出高度的適應性。

為進一步提升內眼角檢測的準確性，一項研究提出了一種高精度的內眼角檢測演算法 [9]，並將其應用於眼球中心定位過程。如圖 2.4 所示，該演算法通過五階段的逐步優化流程，從粗略的眼角區域定位開始，逐步縮小範圍，最終通過梯度信息與形狀約束確定精確的內眼角坐標。該圖展示了內眼角特徵的可視化效果，包括初始定位與最終結果。該方法在正常光線條件下表現優異，且在佩戴透明眼鏡的場景中能抵抗光線反射干擾，顯著提高了眼角檢測的穩定性。研究結果表明，將此檢測算法與三維眼球模型相結合，有助於進一步提升視線追蹤系統在不同場景下的整體性能。



Figure 2.4: 更精準的眼角影像

內眼角與眼球中心的幾何關係可以用數學模型進行精確描述。內眼角被用作參考點，其相對位置由以下數學公式表示：

$$F = f(d_F, \theta_F, \phi_F) = d_F R_z(\phi_F) R_y(\theta_F) \vec{k} + E_c \quad (2.6)$$

公式中， d_F 表示內眼角到眼球中心的距離， θ_F 和 ϕ_F 分別為極角與方位角，用於描述眼球中心在空間中的方向。旋轉矩陣 $R_z(\phi_F)$ 和 $R_y(\theta_F)$ 負責模擬眼球在三維空間中的旋轉，向量 $\vec{k} = [0, 0, 1]^T$ 定義了標準方向，而 $E_c = [x_c, y_c, z_c]^T$ 則表示內眼角的絕對坐標。此公式構建了內眼角和眼球中心之間的精確幾何關係，如圖 2.5 所示，並提供了對動態場景中的頭部運動進行補償的數學基礎。

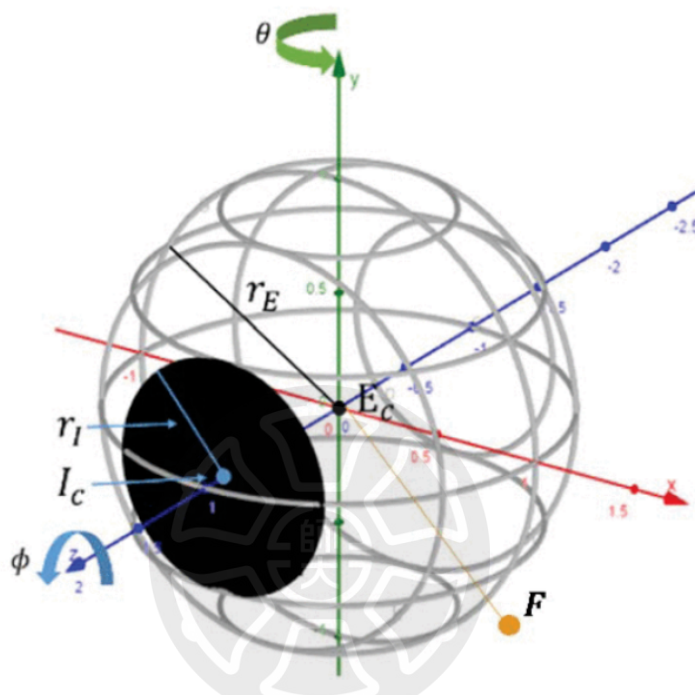


Figure 2.5: 內眼角定位眼球模型的示意圖，展示了內眼角與眼球幾何中心的相對位置及模型匹配過程。

該模型應用在動態場景中（例如虛擬現實或擴增實境應用）時，能夠通過旋轉矩陣動態調整眼球中心位置，從而實現高效的視線補償與追蹤。這不僅提升了系統的準確性，也在頭部運動頻繁或光照條件多變的場景中展現出卓越的穩定性。

然而，基於眼角的定位方法在某些特殊情境下仍可能面臨挑戰。例如，佩戴有色眼鏡時，眼鏡框架可能遮擋部分眼角特徵，導致檢測偏差。此外，在光線不足或強烈背光的情況下，影像對比度下降，使得眼角邊緣的可見性降低，進一步影響系統的檢測精度。針對這些問題，未來研究可以結合更強的抗干擾技術，例如利用深度學習模型對眼角區域進行多模態融合檢測，或者採用紅外與 RGB 影

像的混合數據來提升模型的穩健性與泛化能力。這些改進有望進一步提升視線追蹤系統在多樣化應用場景中的性能。

2.3 頭部姿態估計

傳統基於可見光的眼動儀技術在頭部轉動時容易受到干擾，導致凝視點測量的準確性顯著下降。這種情況在光學路徑變化和幾何投影失真等問題下尤為突出，因此，從單張影像中準確預測頭部姿態並對凝視點進行補償成為了一項關鍵研究任務。如圖 2.6 所示，頭部姿態估計作為解決此問題的核心技術，通過測量頭部的俯仰角 (pitch)、偏航角 (yaw) 和翻滾角 (roll) 來修正凝視點位置，能有效提升眼動儀的精度。同時，這項技術在虛擬現實、駕駛員監控、醫療輔助診斷等領域也具有重要應用價值

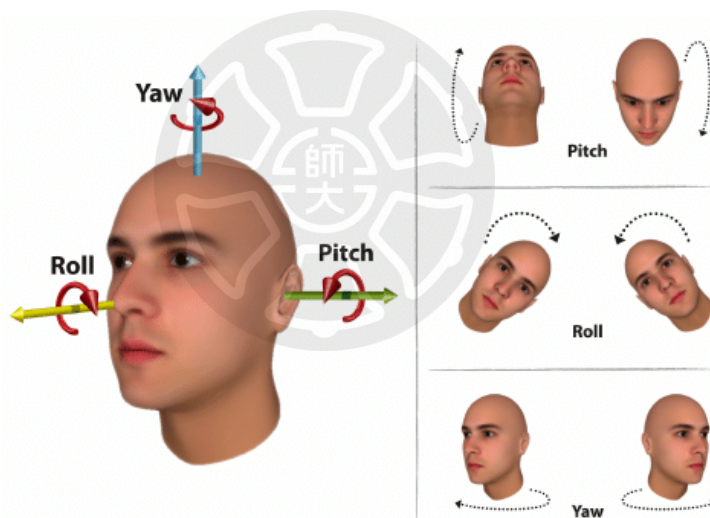


Figure 2.6: 頭部在三維空間中的旋轉方向 [2]

目前，頭部姿態估計的研究主要分為三大類技術方法：基於幾何特徵的方法、基於深度學習的方法和基於三維模型的方法。

2.3.1 基於幾何特徵的方法

基於幾何特徵的方法是頭部姿態估計的早期主流技術，其通過提取面部標誌點（例如眼睛、鼻子和嘴巴）並結合已知的三維頭部模型進行計算。該類方法常利用 PnP（Perspective-n-Point）演算法計算頭部姿態，並使用 RANSAC（隨機抽

樣一致性) 進一步增強對噪聲的魯棒性. 這些方法具有計算效率高的優勢, 並且在低算力環境下表現出色。然而, 幾何特徵方法對光照變化和遮擋情況十分敏感。例如, 在光線不足的情況下, 面部特徵點的提取可能不準確, 進而影響姿態估計的精度。此外, 該方法對頭部大角度旋轉的適應能力有限, 當標誌點被部分遮擋時, 計算結果往往不可靠。

2.3.2 基於深度學習的方法

基於深度學習的頭部姿態估計方法通過訓練卷積神經網絡 (CNN) 從圖像中直接回歸姿態角度, 或者對角度進行分類處理。與基於幾何的方法相比, 深度學習技術克服了光照變化和遮擋的缺陷, 且對頭部大角度旋轉具有更好的魯棒性。例如, Ruiz et al. 提出的 Hopenet 模型利用 ResNet 為基礎網絡結構, 進行回歸與分類相結合的姿態估計, 實現了較高的準確度 [3], 如圖 2.7。同時, Valle et al. [10] 探討了多任務學習框架, 將姿態估計與面部關鍵點檢測結合, 通過共享網絡參數進一步提升模型性能。

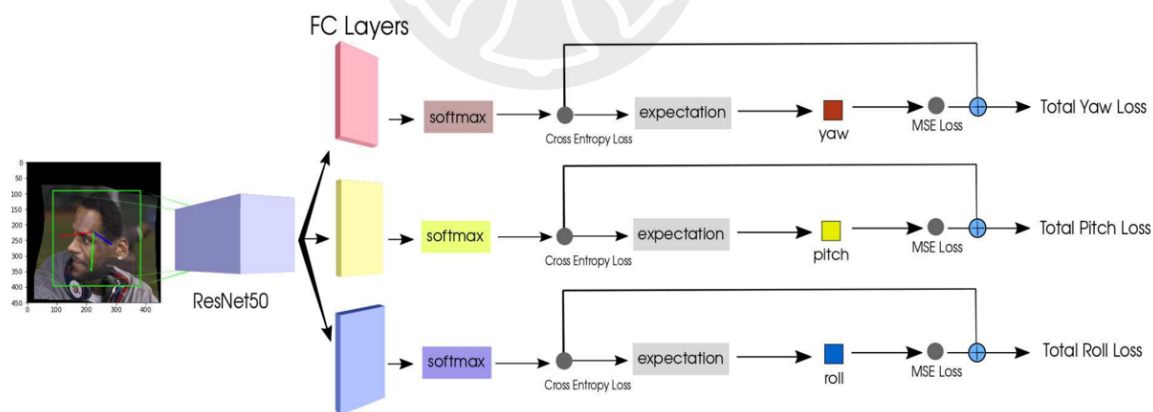


Figure 2.7: 基於 ResNet 的頭部姿態估計架構 [3]

然而, 深度學習方法在頭部姿態估計中也存在一些挑戰和限制。首先, 這類方法高度依賴於大規模的標註數據集, 而這些數據的收集與標註成本往往較高。此外, 深度模型的計算資源需求較大, 例如 ResNet 結構需要高效能硬體支持 (如 GPU 或 TPU), 對於資源有限的環境可能並不友好。

此外，深度學習模型在應對特定問題時仍有改進空間。對於小角度頭部變化的姿態估計，回歸模型可能表現出不穩定，特別是當訓練數據分布不平衡時。此外，許多深度學習方法依賴於預訓練模型（例如 ResNet 的 ImageNet 預訓練權重），當應用場景數據分布與預訓練數據集差異過大時，可能導致性能下降。因此，未來的研究需在數據資源利用、計算效率及模型穩健性上進一步探索，才能更好地發揮深度學習在頭部姿態估計中的潛力。

2.3.3 基於三維模型的方法

基於三維模型的方法通過構建高精度的三維頭部模型，並匹配影像中的特徵點來推算姿態，通常能實現極高的精度。例如，基於 Morphable Model 的方法通過生成三維頭部模型，並使用優化算法使模型與影像中的頭部特徵對齊，以計算姿態角度。Blanz 和 Vetter [4] 的 3D Morphable Model (3DMM)，如圖2.8，是該領域的奠基性研究，並在後續工作中被廣泛應用。



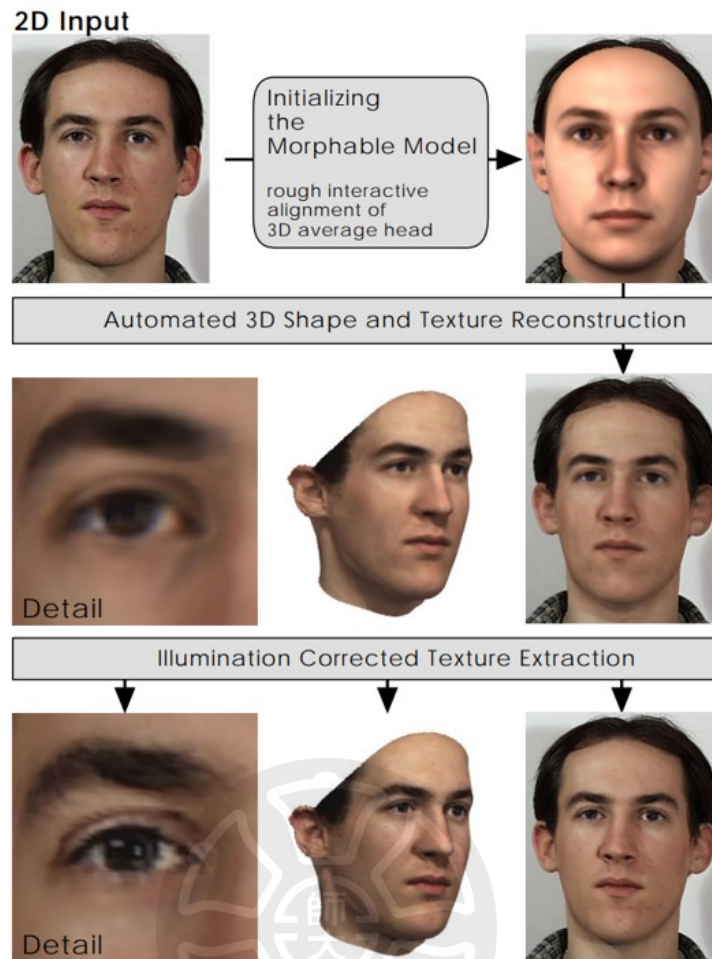


Figure 2.8: 重建 3D 頭部形狀和紋理的處理步驟 [4]

此外，近期研究還結合深度學習與三維模型技術，提出了基於神經渲染的姿態估計方法，例如 DECA (Detailed Expression Capture and Analysis) 模型通過深度網絡回歸三維頭部幾何參數，進一步提升了對姿態和面部表情的重建精度 [11]。然而，這類方法計算代價較高，對於實時應用的適用性有限，更適合需要高精度的場景，如醫療影像分析或高端虛擬角色建模。

2.4 虹膜分割模型

虹膜分割模型是視線追蹤系統的關鍵組成部分，其目標是準確分割眼睛影像中的虹膜邊界，為後續的虹膜匹配與視線估計提供基礎。這些模型在虛擬現實、增強現實、人機交互以及醫學影像處理等多個領域中發揮著重要作用。在可見光和不可見光（紅外光）場景下，分割模型需要適配不同的影像特性與數據集特

徵，以確保在各種條件下都能提供穩定的性能。

Ritnet[5] 是一種專為實時應用設計的虹膜分割模型，其基於結合 U-Net [12] 和 DenseNet[13] 的深度學習技術實現高效的語義分割。該模型大小僅為 0.98 MB，通過五層 Down-Blocks 和四層 Up-Blocks 的架構設計，配合跳躍連接 (skip connections)，在保證計算效率的同時提升了分割精度。RITnet 在 OpenEDS[14] 數據集上的測試表明，它能在不同光線條件下保持良好的性能，並在 NVIDIA 1080 Ti 上達到超過每秒 300 幀的實時處理速度。此外，該模型結合固定伽馬校正和 CLAHE 等預處理技術，有效應對影像對比度差異，顯著提高了虹膜和瞳孔區域的分割準確性。

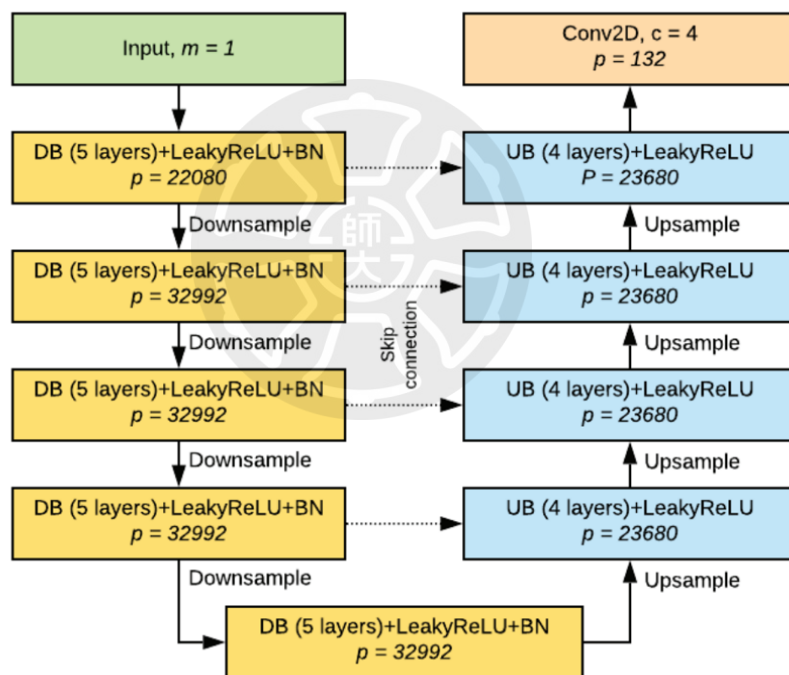


Figure 2.9: RITnet[5] 的架構細節與數據集影像差異示例。

如圖 2.9 所示，RITnet 的網絡架構中採用了五層 Down-Blocks 和四層 Up-Blocks，並通過跳躍連接有效融合多層特徵表示，實現了對虹膜邊界的準確分割。此外，該圖也展示了 IR sensor 和非 IR sensor 數據集的影像特徵差異：IR sensor 數據集的影像通常具有高對比度，並減少了背景干擾；相比之下，RGB 數據集受到光照變化的顯著影響，可能包含陰影、反射和複雜的背景紋理。這些

區域，實現更精確的互動設計與分析。這一過程涉及多種因素的綜合考量，包括視線追蹤技術的精度、螢幕幾何特性、使用者的頭部位置以及眼球運動的生理特性。透過建立數學模型，這些因素得以量化並融入公式設計，使得映射結果能適應不同場景與需求的變化，為實際應用提供理論支持和技術保障。

2.5.1 映射關係建立方式

在視線追蹤系統中，為了將使用者的視線方向（以水平角度 θ 和垂直角度 ϕ 表示）映射到螢幕上的座標點 (x, y) ，通常會使用回歸分析來建立映射模型。在校正過程中，系統要求使用者注視螢幕上的九個校正點，這些點分佈於螢幕的中心、四個角落以及四條邊的中點位置，形成三行三列的佈局（如圖 2.11 所示）。這種佈局能夠獲取視線在螢幕不同區域的數據，為後續回歸模型的訓練提供充分的資料。

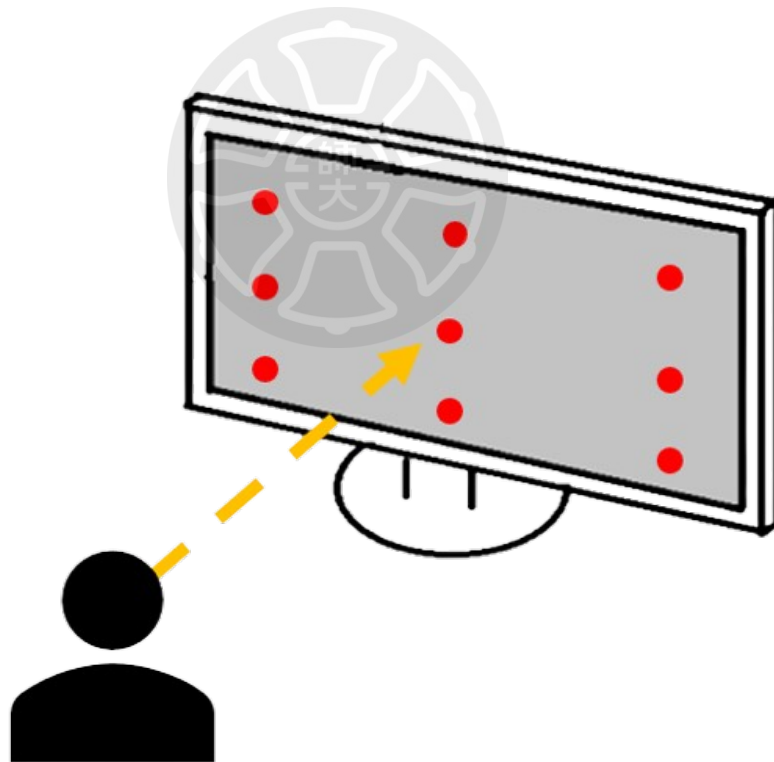


Figure 2.11: 使用者觀看螢幕九點示意圖，

2.5.2 回歸分析

透過回歸分析，視線方向與螢幕座標之間的關係可以被數學模型準確表達為以下形式：

$$\begin{aligned}x &= f(\theta, \phi) + \epsilon_x \\y &= g(\theta, \phi) + \epsilon_y\end{aligned}\tag{2.7}$$

其中， θ 和 ϕ 表示水平和垂直視線角度， f 和 g 是待估計的映射函數， ϵ_x 和 ϵ_y 是誤差項。回歸分析的核心在於通過校正數據估計出這些映射函數，從而建立視線與螢幕座標之間的準確映射。

在視線追蹤中，常用的回歸方法包括線性回歸和非線性回歸（如多項式回歸）。兩者在實現視線映射時各有其優缺點。線性回歸假設視線角度與螢幕座標之間的關係是線性的，其優點是簡單易用、計算效率高且適合小數據集，對於視線變化規律較簡單的場景效果良好。然而，其缺點在於無法捕捉非線性關係，導致模型準確性不足，並且對數據中的異常點較為敏感。

相比之下，非線性回歸（如多項式回歸）允許模型捕捉更複雜的非線性映射關係，適合用於視線與螢幕之間的關係較為複雜的場景。其優勢在於更高的靈活性與準確性，但也帶來了較高的計算成本和過擬合風險。此外，非線性回歸通常需要更多的校正數據來保證模型的穩定性，這在實際應用中可能增加系統的設計成本。

以多項式回歸為例，其數學形式為：

$$\begin{aligned}x &= \beta_0 + \beta_1\theta + \beta_2\phi + \beta_3\theta^2 + \beta_4\theta\phi + \beta_5\phi^2 + \dots + \epsilon_x \\y &= \alpha_0 + \alpha_1\theta + \alpha_2\phi + \alpha_3\theta^2 + \alpha_4\theta\phi + \alpha_5\phi^2 + \dots + \epsilon_y\end{aligned}\tag{2.8}$$

其中， $\beta_0, \beta_1, \beta_2, \dots$ 和 $\alpha_0, \alpha_1, \alpha_2, \dots$ 是待估計的回歸係數， ϵ_x 和 ϵ_y 是誤差項。

在實際應用中，根據具體需求選擇適當的回歸方法十分重要。如果場景對精

度要求較低且追求高效計算，線性回歸是更合適的選擇；若需要更高的準確性並且能夠承擔較高的計算成本，非線性回歸則能提供更優的結果。通過綜合考慮這些因素，可以實現視線追蹤系統在準確性和效率之間的平衡，為各類人機交互場景提供可靠的技術支撐。

2.5.3 偵錯點方法

在視線映射過程中，由於使用者的注視行為可能受到多種因素的影響，例如頭部移動、視線偏移或硬體誤差，部分校正點的數據可能會出現偏差。如果不對這些偏差點進行處理，將它們納入映射公式中，可能會導致模型的準確性下降，最終影響視線追蹤系統的穩定性和性能。因此，對錯誤點的偵測與排除是一個關鍵步驟，能有效提高映射公式的準確性和可靠性。

針對上述問題，錯誤點方法應運而生。其核心在於利用三個穩定的視點來建立初步的回歸公式，並以此公式對其他視點進行判斷與調整。通過這一方式，可以有效找出錯誤的視點，並重新校正以提升整體的映射準確性和穩定性。

操作流程具體如下：首先，從所有校正點中選擇三個穩定的視點作為正確點，用以構建初步的回歸公式。選擇這三個視點的原則是確保其具有最佳的空間分佈，例如在螢幕中央和對角線附近，以提高公式的穩定性和可靠性。隨後，利用建立的回歸公式，計算其他視點相對於公式的偏差。如果偏差值（例如角度或距離）較小，則該點被認為是映射準確的正常點；反之，偏差較大的視點可能存在異常，需要進一步分析。

接下來，計算每個視點的「偏差成本」(cost)，這是一個衡量視點偏差程度的指標，通常以角度偏差為基準，並與預定的誤差容忍範圍（例如 2.5° ）進行比較。當發現偏差超出容忍範圍的視點時，這些視點會被暫時標記為異常點，並被排除在公式之外。接著，重新利用穩定的視點來修正回歸公式，這一過程反覆進行，直到公式能準確覆蓋所有視點並達到預期精度。

錯誤點方法的主要特點包括以下幾點：首先，該方法不需要一次性處理所有視點，只需挑選三個穩定的視點即可啟動公式建構，從而大幅降低了初期的計算複雜性。其次，當異常點被識別後，可以通過排除異常點來不斷改進公式，最終提升整體映射精度。最後，該方法允許根據實際需求調整誤差容忍範圍，或者選擇更多的視點參與校正，從而適應不同的應用場景。

在實際應用中，使用者通常需要注視螢幕上的九個校正點，記錄其視線分布數據。然而，由於視線追蹤設備的精度限制或使用者的注視偏移，可能會出現某些點偏差較大的情況。例如，使用者在注視螢幕四個角落的點時，由於頭部角度的變化，數據偏差可能顯著增大，如圖]2.12。這些偏差點即為錯誤點的典型例子，若不加以排除，可能會導致整體回歸公式的準確性下降。通過錯誤點方法，可以有效找出這些偏差較大的點並將其排除，從而提升模型的穩定性和準確性。

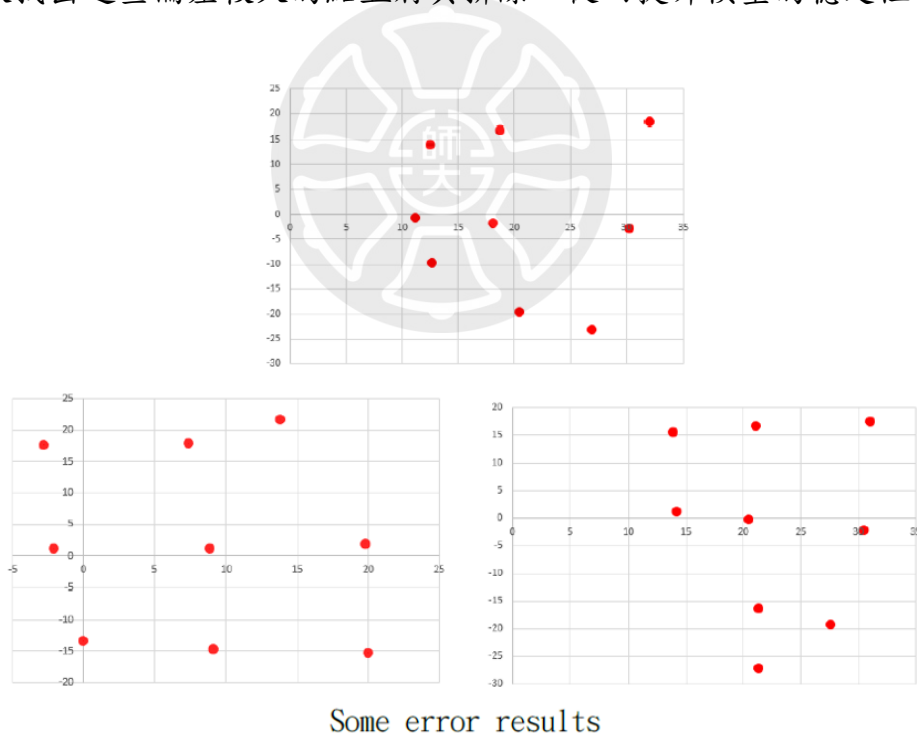


Figure 2.12: 不同使用者觀看螢幕九點視線分布，可以觀察到會有錯誤點的產生。

第三章 系統架構與設計

本篇章將詳細介紹本研究所開發的系統架構，並全面闡述各模組的設計理念與實現細節。整體系統的架構概覽如 Figure 3.1 所示。首先，在章節3.1中，我們採用了 Mediapipe 提供的臉部特徵點偵測算法，以高精度地定位影像中的臉部特徵點，為後續處理提供穩固的基礎。在章節3.2中，我們將進一步描述如何利用現有的頭部姿態估計技術，動態更新每一幀臉部模型在三維空間中的位置。同時，根據不同使用者的特徵，自適應調整其 3D 頭部模型，使系統能夠靈活適應各種使用者。這些模組的結合確保了整體系統的穩定性與普適性。接下來，在章節3.3中，我們將說明如何基於現有的可見光眼睛影像資料集，訓練一個專門的虹膜分割模型。該模型旨在提升系統的穩健性與準確性，從而更加精確地提取與分析虹膜的相關參數。在章節3.4部分，我們將詳述如何結合已構建的 3D 眼球模型與影像特徵，通過最佳化方法推導眼球旋轉角度 θ 與 ϕ 的最優解，進一步增強系統的空間定位能力。最後，章節3.5，我們將探討如何通過建立凝視方向與螢幕之間的映射關係，實現高效的人機互動功能。

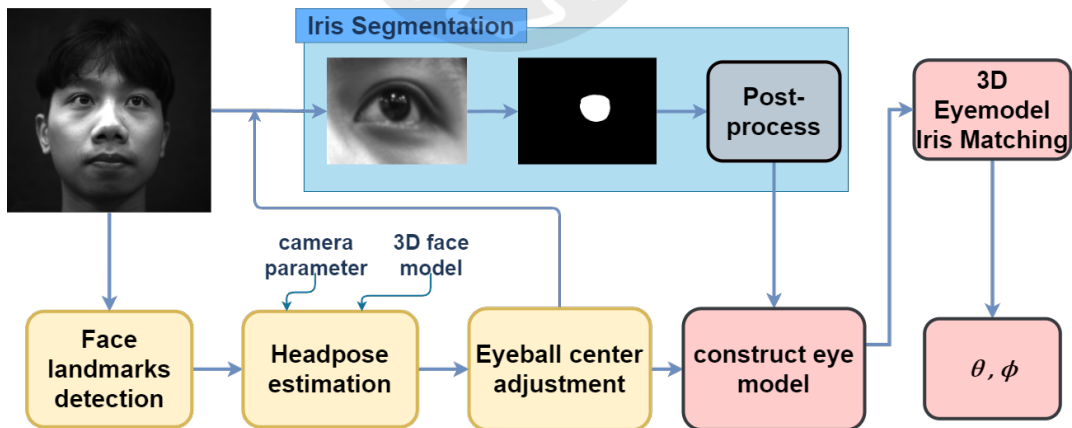


Figure 3.1: 系統架構圖

3.1 臉部特徵點的定位與提取

臉部特徵點的準確定位是整個系統的基礎，直接影響後續模組的表現。因此，在本研究中，我們採用了 Mediapipe 提供的臉部特徵點偵測算法。該算法以其高效性和穩健性著稱，能夠在多種光照條件和不同使用者面貌特徵下，實現精

準的特徵點檢測。

在系統實現中，我們首先對輸入的臉部影像進行預處理，包含灰階化與高斯模糊兩個步驟，這些處理對於視線追蹤任務的準確性與穩定性至關重要。首先，灰階化將彩色影像轉換為單通道的灰階影像，有效降低數據的維度，從而減少計算負擔並提升處理效率，這對於需要實時運行的視線追蹤系統尤為重要。同時，灰階化能去除影像中的色彩信息，避免光照變化或膚色差異對視線追蹤的干擾，突出眼部結構的幾何特徵，如瞳孔與虹膜邊界。此外，灰階化還增強了影像中關鍵區域的對比度，使得眼部特徵在後續處理中更容易被檢測和提取。

在此基礎上，我們進一步對灰階影像應用高斯模糊處理。高斯模糊能有效平滑影像，抑制由光線、相機噪聲或其他環境因素引起的隨機干擾，從而確保視線追蹤過程中的穩定性。模糊處理還能提升邊緣檢測的穩定性，特別是在提取瞳孔和虹膜邊界時，能過濾掉多餘的紋理和雜訊，集中於影像的關鍵特徵區域。此外，高斯模糊的平滑效果還有助於特徵提取算法的準確性，確保在高光或低光場景下，眼部特徵仍能被精確捕捉，進一步提升視線追蹤的整體性能。

接著，利用 Mediapipe 的 Face Mesh 模組進行特徵點偵測，並提取出關鍵的特徵點作為後續模組的輸入。該模組能生成 468 個臉部特徵點，如 Figure 3.2，覆蓋臉部的主要區域，提供了豐富的幾何信息。相比傳統方法（如 Dlib 或 OpenCV），Mediapipe 的檢測精度和運算效率更為出色，特別適用於本研究所需的即時應用場景。



Figure 3.2: mediapipe 臉部特徵點

在提取特徵點後，為了進一步提升系統的穩定性和抗噪能力，我們對一定幀數量的特徵點進行了濾波處理。具體而言，我們採用了移動平均濾波器（Moving Average Filter）對多幀的特徵點進行平滑，以減少由於檢測波動或環境噪聲引起的不穩定性。該方法通過計算特徵點在一定時間窗口內的平均值，有效平滑了因光線變化、頭部微動或檢測模型本身波動帶來的數據偏差。

此外，這種濾波策略還能提升系統的抗干擾能力，特別是在高動態場景中，例如使用者快速轉頭或面部表情變化時，特徵點的位置可能出現瞬時的異常波動。通過濾波處理，這些波動被有效抑制，確保系統輸出的特徵點數據更加穩定，進而為後續的視線追蹤模組提供更高質量的輸入。

總之，對特徵點進行濾波不僅能提升整體檢測的穩定性，還能增強系統在動態環境下的適應性，為實現高準確度的即時視線追蹤提供了重要的技術保障。這一步驟與 Mediapipe 提供的高效特徵點檢測相結合，使本系統能在複雜環境下保持卓越的性能表現。

3.2 基於頭部姿態的眼球中心校正

眼球中心的準確定位是視線估計和人機交互等應用中的關鍵步驟。然而，由於頭部的姿態變化會導致眼球位置的視覺偏差，直接影響系統的穩定性與精準性，因此需要對眼球中心進行基於頭部姿態的校正。

在多數應用場景中，使用者的頭部可能以不同角度旋轉或偏移，這會改變相機捕捉到的眼球中心相對位置。例如，在頭部偏轉時，未經校正的眼球中心可能偏離實際位置，進而影響下游模組的計算準確性。因此，校正算法需要根據頭部姿態動態調整眼球位置，確保其與真實位置一致。

本研究使用章節3.1 所提到的 Mediapipe 的 Face Mesh 模組進行頭部姿態估計。該模組提供臉部特徵點座標，結合三維幾何模型，可以計算頭部在空間中的旋轉和平移參數（Pitch, Yaw, Roll）。為了提升估算的精度，我們參考了 PnP（Perspective-n-Point）算法來進行優化，利用臉部特徵點與三維模型的對應關係計算出精確的姿態參數。

本研究中，我們建立了一個基於頭部姿態的校正模型，具體步驟如下：

3.2.1 頭部姿態估算技術

頭部姿態估算是眼球中心校正過程中的關鍵步驟，旨在獲取頭部在三維空間中的旋轉角度（Pitch、Yaw、Roll）和平移位置，從而補償由頭部姿態變化引起的視覺偏差，確保系統穩定且高精度輸出。為實現這一目標，本研究基於 PnP（Perspective-n-Point）演算法計算頭部姿態，並引入 Fornalczyk et al. [16] 提出的動態頭部模型優化方法，針對初始三維頭部模型進行迭代優化，以提升姿態估算的準確性。

首先，從 Mediapipe Face Mesh 模組中提取臉部的 468 個特徵點，這些點覆蓋了眉毛、眼睛、鼻子、嘴巴、下巴等關鍵部位。本研究從中篩選出 14 個穩定的關鍵點，包括眉毛與雙眼的內外角、鼻尖與鼻翼兩側、嘴角及下巴中心，如

圖3.3。這些點的位置在頭部運動時具有穩定性，適合與三維頭部模型進行匹配。初始三維頭部模型則基於標準人頭尺寸，提供標準化的三維坐標，作為後續姿態估算的輸入。



Figure 3.3: 篩選出來的 14 個臉部特徵點。

基於透視投影原理，頭部姿態估算遵循以下數學模型：

$$\mathbf{p}_i = \mathbf{K} \cdot (\mathbf{R} \cdot \mathbf{P}_i + \mathbf{T}), \quad (3.1)$$

其中， \mathbf{p}_i 是圖像平面中的二維特徵點， \mathbf{P}_i 是三維頭部模型中的對應點， \mathbf{K} 是相機內參矩陣， R 和 T 分別表示頭部的旋轉矩陣和平移向量，可以參考 fig 3.4 的示意圖。為最小化觀測到的二維特徵點與三維模型投影點之間的誤差，我們使用 **Levenberg-Marquardt 優化演算法** 來求解頭部姿態參數 (R 和 T)。然而，由於初始三維頭部模型基於平均人頭尺寸，並未完全符合個體特徵，可能會產生一定的估算誤差。

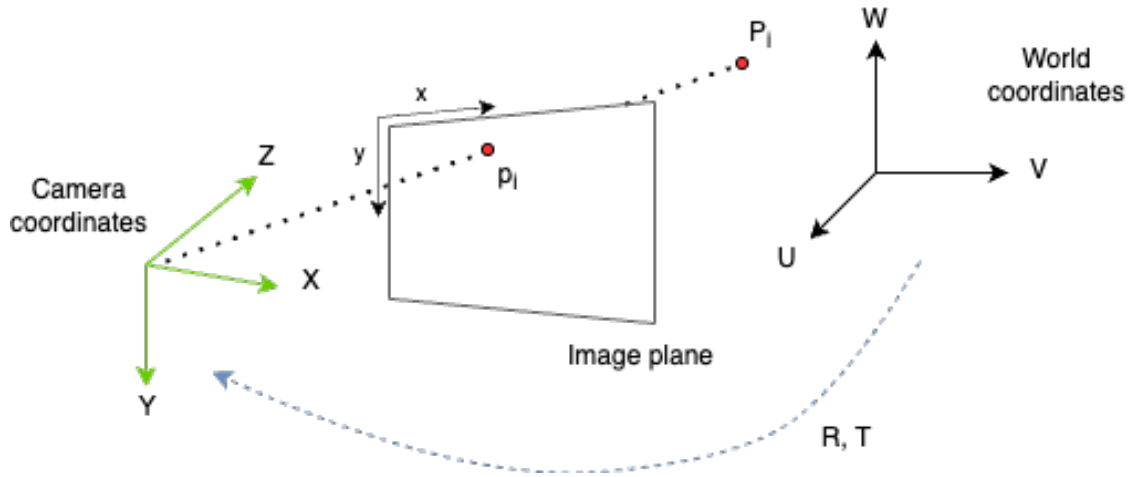


Figure 3.4: PnP 示意圖：3D 點透過相機姿態 (R, T) 投影到 2D 成像平面。

為解決上述問題，我們引入了動態頭部模型優化方法。該方法透過多幀數據進行累積優化，以迭代方式調整三維頭部模型中的點位置，從而提高頭部姿態估算的準確性。具體過程如下：根據每幀圖像的頭部姿態參數 (R 和 T)，我們將三維模型點 \mathbf{P}_i 投影到圖像平面，並計算其重投影點 \mathbf{p}'_i ：

$$\mathbf{p}'_i = \mathbf{K}[R|T]\mathbf{P}_i. \quad (3.2)$$

由於初始模型與真實頭部的差異，重投影點 \mathbf{p}'_i 與實際檢測到的二維特徵點 \mathbf{p}_i 之間會存在一定誤差。我們根據這一誤差計算每個三維點的修正向量 \mathbf{w}_i ，並累積所有幀的修正量後，動態更新三維頭部模型中的點位置：

$$\mathbf{P}_i = \mathbf{P}_i + \frac{1}{N} \sum_{k=1}^N \mathbf{w}_i, \quad (3.3)$$

其中， N 是累積的幀數， \mathbf{w}_i 是針對每個點計算出的修正向量。為了避免頻繁更新引入的噪聲，該優化過程通常在固定的幀間隔（例如每 175 幀）執行一次。

經過動態優化後，三維頭部模型逐漸與使用者的真實頭部特徵相匹配，從而提升頭部姿態估算的準確性。最終，我們輸出優化後的頭部姿態參數，包括旋轉角度 (Pitch、Yaw 和 Roll) 和平移向量 T 。這些參數作為後續眼球中心校正的輸入，能夠有效應對頭部姿態變化，確保系統在實際應用場景中的穩定性和高精度。

輸出。

3.2.2 眼球中心初始計算

眼球中心的準確估算是視線估計和人機交互系統中的重要基礎步驟。本研究基於優化後的三維頭部模型，通過內外眼角的幾何特徵點進行初始眼球中心的計算，並根據頭部幾何特徵向後偏移一個合理的距離，估計出眼球中心的初始位置。

首先，我們從三維頭部模型中提取內眼角和外眼角的三維坐標，分別記為 P_{inner} 和 P_{outer} 。這兩個點是眼部區域最穩定且容易檢測的幾何特徵點，它們的連線可以反映眼睛的水平邊界。根據內外眼角的三維坐標，我們計算兩點之間的中點，作為眼球中心的初始水平位置，具體計算公式如下：

$$P_{mid} = \frac{P_{inner} + P_{outer}}{2}, \quad (3.4)$$

其中 P_{mid} 表示內外眼角的幾何中點。

該中點位於頭部表面平面上，但實際眼球的中心應該位於頭部內部，因此需要進行後向偏移。為了確定後向偏移的距離，我們首先計算內外眼角之間的歐幾里得距離 d ，表示為：

$$d = \|P_{outer} - P_{inner}\|. \quad (3.5)$$

接著，將該距離除以 2，並乘上一個預設的尺度因子 α （表示眼球半徑的比例關係），作為眼球中心相對於中點向後偏移的量。偏移方向設置為頭部坐標系中的 $(0, 0, 1)$ ，即沿著 Z 軸負向朝頭部內部平移，得到眼球中心的最終估計位置 P_{eye} ：

$$P_{eye} = P_{mid} - \alpha \cdot \frac{d}{2} \cdot (0, 0, 1), \quad (3.6)$$

其中 α 是一個可調整的比例因子，用於適配不同個體的眼球半徑大小，並根據實驗數據進行合理設置。

這樣的設計有以下幾點優勢：首先，內外眼角是穩定的幾何特徵點，易於檢測，並且能很好地反映眼睛的邊界特徵。其次，基於眼角間距與尺度因子的後向偏移，直接利用了頭部坐標系的 Z 軸方向，避免了過於複雜的法向計算，且具有較高的計算效率。最後，尺度因子的靈活性允許根據不同個體的頭部和眼球結構進行微調，提升結果的準確性與方法的適用性。

通過上述方法計算得到的眼球中心，雖然是初始估計值，但為後續基於頭部姿態的動態校正提供了一個穩健且合理的初始位置，有助於提升視線估計的整體精度與系統穩定性。

3.2.3 動態校正方法

在估算出頭部姿態參數 $[R|T]$ 以及初始眼球中心位置後，為了適應連續影像中頭部姿態的動態變化，本研究提出了一種動態更新眼球中心並將其投影到影像平面的方法。該方法能夠根據每一幀的頭部姿態參數，重新計算眼球中心的位置，並確保其在不同頭部運動下的準確性與穩定性。

首先，基於頭部姿態估算結果，三維旋轉矩陣 R 和平移向量 T 描述了頭部從模型坐標系到相機坐標系的姿態變換。我們利用這一變換，將初始眼球中心 \mathbf{P}_{eye} 位置動態更新到相機坐標系中，具體公式如下：

$$\mathbf{P}_{\text{eye, cam}} = R \cdot \mathbf{P}_{\text{eye}} + T, \quad (3.7)$$

其中 $\mathbf{P}_{\text{eye, cam}}$ 是當前幀中眼球中心在相機坐標系中的三維坐標， R 和 T 分別是當前頭部姿態的旋轉矩陣與平移向量， \mathbf{P}_{eye} 是靜態頭部模型中的初始眼球中心位置。

接下來，為了將更新後的三維眼球中心位置映射到影像平面上，我們基於透視投影模型，利用相機內參矩陣 \mathbf{K} 進行投影。相機內參矩陣 \mathbf{K} 包含焦距 f_x 、 f_y

和光學中心 (c_x, c_y) ，具體形式為：

$$\mathbf{K} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}. \quad (3.8)$$

透過相機內參矩陣，我們將三維坐標 $\mathbf{P}_{\text{eye, cam}}$ 投影到影像平面上，得到二維像素坐標 \mathbf{p}_{eye} ：

$$\mathbf{p}_{\text{eye}} = \mathbf{K} \cdot \mathbf{P}_{\text{eye, cam}}. \quad (3.9)$$

其中 $\mathbf{p}_{\text{eye}} = (u, v)$ 表示眼球中心在當前影像平面中的二維坐標，整體的投影關係可以參考圖3.5。

總結來說，通過頭部姿態變換和透視投影，本方法能夠根據每一幀影像的頭部姿態參數動態更新眼球中心位置，並將其準確地映射到影像平面上。這種方法既簡單高效，又能適應頭部運動過程中的姿態變化，為視線估計和人機交互系統提供了連續且穩定的眼球中心輸出。

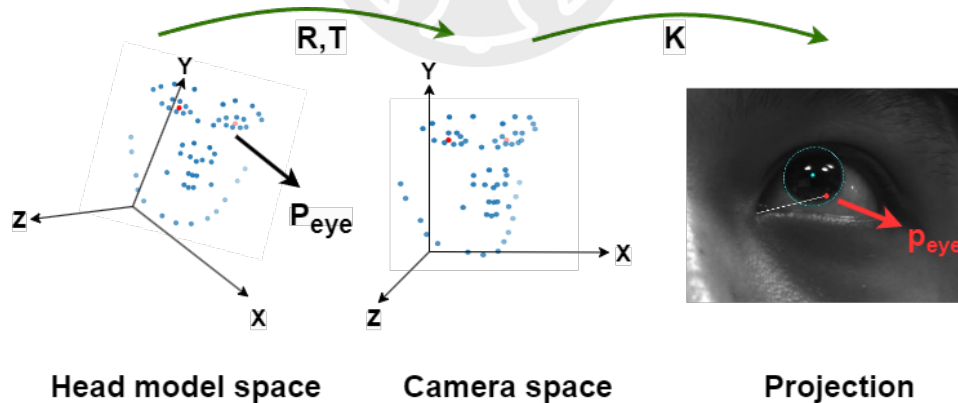


Figure 3.5: 眼球位置更新的流程：通過旋轉 R 和平移 T 從頭部模型空間轉到相機空間，再經由 K 映射至圖像空間，最終確定眼球中心 P_{eye} 的位置。

3.3 虹膜區域分割

虹膜區域分割是視線估計與生物識別中至關重要的一步。虹膜作為眼球中穩定且唯一的紋理結構，具有高區分度與穩健性。然而，由於眼部圖像可能受光

照、遮擋、頭部運動和反射等因素影響，準確分割虹膜區域並提取相關參數具有一定挑戰。本研究從虹膜區域的偵測、分割後的數據清理到參數提取，提出了一套完整的虹膜分割處理流程。

為了精確定位虹膜區域，本研究基於輸入的眼部圖像，使用多階段處理方法來完成虹膜區域的偵測與粗定位。具體流程如下：

3.3.1 虹膜區域的偵測方法

根據參考的論文 RITnet [5] 的框架，我對模型進行了修改，將訓練數據中的標註 (ground truth) 調整為僅包含虹膜區域與背景。因此，模型的輸出通道數由原本的 $w \times h \times 4$ (對應四個分類：背景、虹膜、鞏膜和瞳孔) 減少為 $w \times h \times 2$ (僅對應虹膜和背景)。隨後，我基於這一簡化的分割任務對模型進行了重新訓練。

這一修改的目的是聚焦於虹膜區域的分割，並簡化網路的輸出結構，以提升在特定應用場景中的效能，特別是在不需要對其他眼部區域 (例如鞏膜或瞳孔) 進行詳細分割時。這一調整同時也使模型的計算效率得到了進一步的優化。在某些應用場景中，例如基於虹膜的目光追蹤或身份識別，虹膜是主要的關鍵特徵，而其他眼部區域的詳細分割 (如鞏膜和瞳孔) 對性能的影響相對有限。將分割任務簡化為僅處理虹膜與背景，不僅可以更高效地利用模型資源，還能縮短推論時間，使模型在資源受限的環境中更加實用。

此外，本研究的技術選擇還考慮了兩個重要因素。第一，論文中所使用的數據來自於紅外 (IR) 相機拍攝，而我們的研究使用的是可見光相機。紅外相機拍攝的圖像可以更明顯地區分瞳孔與虹膜，因為瞳孔通常在紅外光譜下呈現非常深的黑色，而虹膜則有顯著的灰階差異。然而，在可見光條件下，瞳孔與虹膜的顏色差異減弱，使得分割的準確性面臨挑戰。因此，我們採用將標註簡化為虹膜與背景的方法，減少由於瞳孔與虹膜區分模糊而引起的誤差。

第二，這一方法也適應了亞洲人眼睛的特徵。亞洲人的瞳孔與虹膜顏色更為接近，使得虹膜和瞳孔的邊界在可見光相機下更難以區分。聚焦於虹膜與背景的分割，能減少標註不準確對模型性能的干擾，同時也更適應亞洲人眼睛的特徵。原始標註數據中的鞏膜和瞳孔邊界可能因標註質量不一或主觀判定的差異而產生偏差，而聚焦於虹膜與背景的分割則能降低這些誤差對模型性能的影響。

最後，輸出通道數的減少有效降低了模型的計算複雜度，進一步提升了推論速度和計算效率。具體的實施細節，包括訓練過程和評估方法，遵循了原始RITnet框架的相關指南。例如，訓練時使用標準的損失函數和學習率調整策略，以確保模型能穩健地收斂。經過這一簡化的設計，模型在效能與資源效率上實現了良好的平衡，為特定應用場景提供了一種高效且準確的解決方案。

在簡化設計的基礎上，模型在效能與資源效率之間實現了良好的平衡，為特定應用場景提供了一種高效且準確的解決方案。為直觀展示這一簡化設計的效果，下圖（圖 3.6）展示了模型在虹膜分割任務中的預測結果。左側為輸入的眼睛圖像，右側為模型分割出的虹膜區域。



Figure 3.6: 左圖為輸入的眼睛影像，右圖為模型輸出的虹膜分割結果。

通過上述圖表可以看出，經過優化後的模型能夠準確地分割出虹膜區域，同時在計算效率和資源利用上實現了良好的平衡。

3.3.2 分割後的數據清理

由於分割模型輸出的影像可能受到環境光線、反射或其他外部因素的影響，導致結果中存在一定程度的雜訊或不完整的區塊，因此需要進一步的數據清理以確保結果的準確性與一致性。針對這些可能的問題，我們採用以下處理流程來優化分割結果。

首先，分割結果中可能出現多個區塊被預測為虹膜區域，而實際上虹膜應該只是一個連續的區塊。因此，我們計算影像中的最大連通區域 (Largest Connected Component, LCC)，並將最大區塊視為真實的虹膜區域。這一方法能有效過濾掉其他不相關的或由雜訊導致的小型區塊。

其次，虹膜區域的形狀應該是凸的幾何形狀，而分割模型的預測結果可能因邊緣模糊或雜訊影響而出現局部的凹陷現象。為了修復這些不規則性，我們採用了凸包演算法（Convex Hull），該方法通過構建包含所有虹膜區域像素的最小凸多邊形來填補凹陷部分，使得結果更加符合虹膜的真實形態。

經過上述處理後，分割結果的完整性和準確性得到了顯著提升，為後續的分析和應用提供了更加穩定的基礎，下圖為處理前後的結果圖3.7。這種清理過程不僅減少了模型在不良環境條件下的誤差影響，還能使分割結果更具可靠性和應用價值。

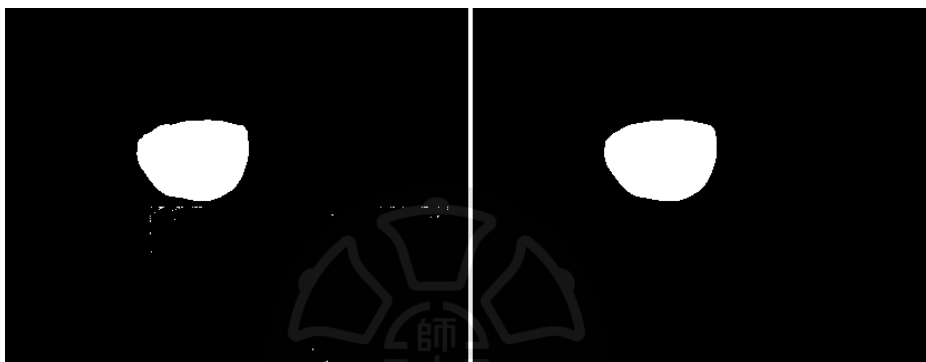


Figure 3.7: 虹膜分割後處理: 左圖為處理前，具有一些雜訊，虹膜輪廓較不清晰; 右圖為處理後，濾掉雜訊且虹膜輪廓較平滑。

3.3.3 虹膜參數提取

為了從影像中準確提取虹膜的半徑和中心位置，我們設計了一套演算法，主要分為兩個步驟。

首先，我們使用最小外接矩形（Minimum Bounding Rectangle）來框住虹膜區域。這一步驟透過邊緣檢測技術（如 Canny 邊緣檢測）提取影像中的邊緣資訊，然後計算包含所有邊緣點的最小矩形，以確定虹膜的大致範圍。這種方法能有效地縮小後續處理的範圍，提高計算效率。

接下來，從該矩形的左右邊緣各取一段線段，分析其像素強度變化，以進一步確定虹膜的半徑和中心位置。具體而言，我們計算這些線段上像素點的平均座標，並以此估計虹膜的中心點。同時，根據線段的長度和位置，推算出虹膜的

半徑。這種方法利用了虹膜在影像中的幾何特徵，能夠提高定位的準確性，附圖3.8為提取虹膜中心與半徑的簡單的流程示意圖。

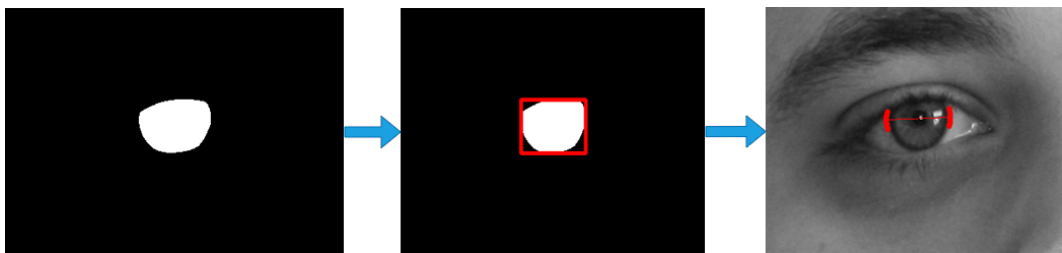


Figure 3.8: 後處理步驟提取虹膜初始中心與半徑。

研究針對虹膜分割提出了一套完整的流程，從影像的預處理、模型設計到後處理的數據清理和參數提取，實現了高效且準確的虹膜區域分割和參數提取方法。

首先，我們透過調整基於 RITnet 的模型結構，專注於虹膜區域與背景的分割，簡化了任務複雜度並提升了計算效率。此模型特別適應可見光拍攝的亞洲人眼睛特徵，有效克服了虹膜與瞳孔邊界模糊的挑戰。

其次，在分割結果的後處理環節，我們結合最大連通區域和凸包演算法，對分割結果進行清理與修正，顯著提高了分割結果的準確性與一致性，並減少了環境雜訊的影響。

最後，通過提取虹膜的中心和半徑等幾何參數，我們的演算法能穩健地生成虹膜相關數據，為視線估計、身份識別等應用奠定了良好的基礎。該方法充分考慮了影像品質與環境條件對性能的影響，並提供了調整和優化的空間。

總而言之，這項研究為可見光環境下的虹膜分割問題提供了一個高效可靠的解決方案，不僅在理論上具有貢獻，也為實際應用場景提供了重要的支持。

3.4 虹膜特徵匹配

在凝視追蹤系統中，凝視方向的微小誤差可能對整體系統性能產生顯著影響。即便是像素級別的偏差，也可能導致視線估計結果的顯著偏移，進而削弱系

系統的穩定性和可靠性。因此，如何確保系統的穩定性和準確性成為了我們研究中的關鍵課題。

在上一章節中，我們提出了一種基於虹膜分割的算法，成功地確定了虹膜的初始半徑以及初始中心位置。這一初始估計值不僅為後續的凝視方向計算提供了可靠的起點，還在一定程度上減少了可能的搜索範圍。然而，為了進一步提高系統的穩定性，我們在此基礎上引入了虹膜匹配技術，旨在基於初始值進行精確的虹膜位置搜尋和調整。這一策略能夠顯著提升凝視追蹤的穩定性和精度，尤其是在動態環境中。

以下內容將我們的研究劃分為若干小節進行詳細討論，從虹膜分割的數據基礎到虹膜匹配的計算過程，逐步探討如何通過優化算法來提升系統的穩定性和實用性。

3.4.1 搜尋範圍優化

在可見光譜視線追蹤系統中，精確計算眼睛的旋轉角度是準確視線估計的關鍵挑戰之一。本研究基於先前提出的 3D 虹膜匹配方法，構建了三維眼睛模型，並結合改進的算法進一步提升了視線追蹤的精度和計算效率。

首先，該方法採用了虹膜匹配技術，通過分析眼球的幾何特徵來確定其旋轉角度。傳統方法需要在所有可能的眼睛旋轉角度範圍內進行搜索，但這種方法往往需要在計算精度與運算負載之間進行權衡，導致系統的實時性能受到限制。例如，傳統方法中， θ 的搜索範圍為 0° 至 30° ，而 ϕ 的搜索範圍則為 0° 至 180° 。這樣大範圍的搜索雖然可以涵蓋多種情境，但計算資源的消耗極高，特別是在需要高搜尋精度的情況下，運算量會隨範圍與步進的縮小而指數增長。依照搜尋精度的設定，例如將步進設定為 0.1° ，需要評估的可能位置數量會大幅增加，進一步加劇運算負擔。因此，如何縮小搜索範圍並優化計算策略，成為提升系統性能的重要挑戰。下圖3.9為虹膜搜索範圍的示意圖。

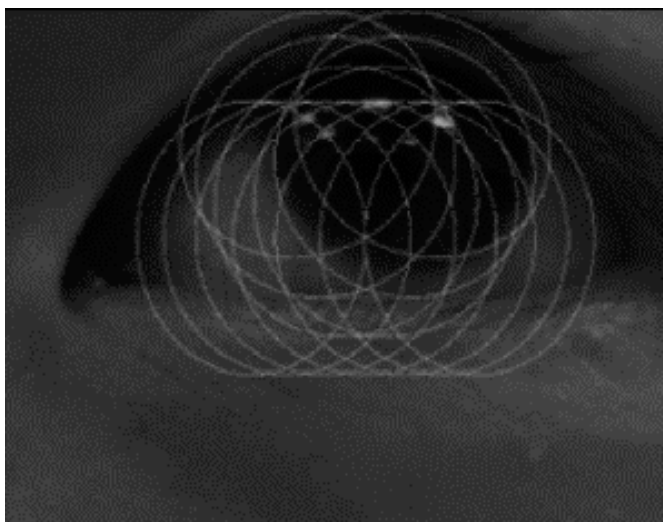


Figure 3.9: 虹膜搜尋範圍示意圖。搜尋範圍越大或搜尋精度縮小，都會導致運算瓶頸。

為了解決這一問題，我們在系統中引入了虹膜分割技術，該技術能夠快速且準確地識別虹膜的初始位置，從而顯著縮小旋轉角度的搜索範圍。在本研究中，我們通過初估旋轉角度的方式，將 θ 的搜索範圍縮小至 θ_0 的 $\pm 5^\circ$ ，而 ϕ 的搜索範圍則縮小至 ϕ_0 的 $\pm 10^\circ$ 。這種方式不僅顯著減少了計算量，還提升了匹配的準確性和效率。

具體而言，該方法通過以下公式計算眼睛的初始旋轉角度 θ 和 ϕ ：

$$\theta_0, \phi_0 = \arctan\left(\frac{\Delta r}{\Delta x}\right) \quad (3.10)$$

其中， Δr 表示虹膜中心與眼球中心在 y 和 z 方向的位移差，而 Δx 表示在 x 方向的位移差。這一公式能夠有效地同時計算繞 z 軸 (θ) 和繞 y 軸 (ϕ) 的旋轉角度。在確定了這些初始角度後，系統進一步在這些角度的鄰域內進行精細搜索，以獲得最佳的匹配結果。

表 3.1 比較了傳統方法與本研究改進方法中， θ 和 ϕ 的搜索範圍。從表中可以看出，改進方法大幅縮小了搜索範圍，有效降低了計算負擔。

相比於傳統的全範圍搜索方法，我們的改進不僅顯著降低了運算複雜度，還有效提升了視線估計的精確度。此技術的核心優勢在於能夠利用虹膜分割結果提

Table 3.1: 旋轉角度搜索範圍比較

方法	θ 範圍	ϕ 範圍
傳統方法	$0^\circ \sim 30^\circ$	$0^\circ \sim 180^\circ$
本研究改進方法	$\theta_0 - 5^\circ \sim \theta_0 + 5^\circ$	$\phi_0 - 10^\circ \sim \phi_0 + 10^\circ$

供的先驗信息，優化後續計算過程，從而實現高效的視線追蹤。

3.4.2 算分區域

由於眼睛在注視不同方向時，虹膜在影像中的形狀與遮擋情況會隨之變化，部分區域甚至可能被眼皮遮住。為了提升凝視方向估計的準確性，本研究基於初始凝視方向，對虹膜邊界進行動態分配和算分處理，以考慮這些不規則變化。

如圖 3.10 所示，當眼睛向左側注視時，虹膜左邊的區域更接近眼皮，該區域的梯度計算值 (gradient) 通常較低，導致其在視線估計中的貢獻減少。基於這一觀察，我們提出了一種根據眼球與眼皮關係，並結合初始凝視方向的邊界算分方式。該方法能夠動態決定虹膜邊界的有效算分區域，並剔除受眼皮遮擋較大的部分，從而提升估計的穩定性與準確性。

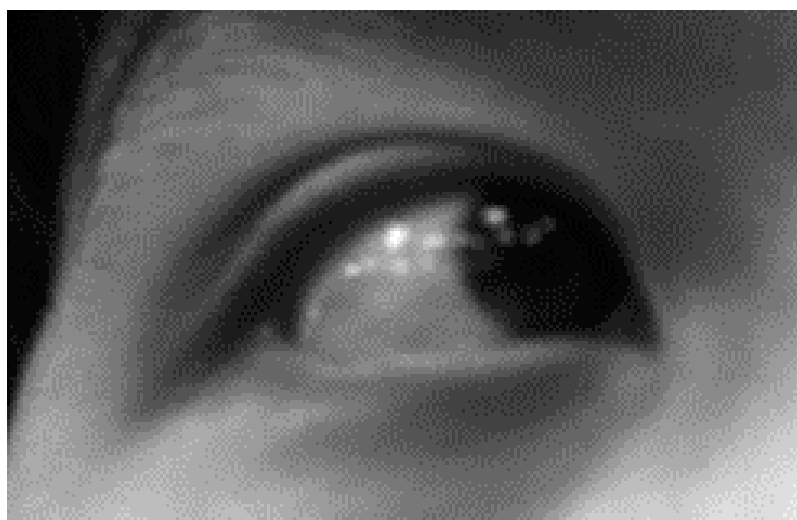


Figure 3.10: 眼睛往左邊看虹膜的與影像上的情況。

在視線追蹤系統的初始階段，我們首先考慮用戶眼睛直視相機的情境。由於

人眼的生理結構以及不同用戶眼睛大小的差異，我們觀察到大部分用戶在直視時並不會完全暴露整個虹膜區域。通常情況下，上下眼皮會遮擋虹膜的部分上下邊界，因此在直視情況下，針對虹膜的完整邊界進行算分並不適用。

為了適應這一特性，我們重新定義了直視情境下的虹膜邊界算分範圍。在這種情況下，虹膜的有效邊界並非整個環形，而是集中於左側和右側的特定區段 (segments)，如圖 3.11 所示。這些有效邊界區段避免了上下眼皮遮擋區域的干擾，從而提升了算分的準確性。

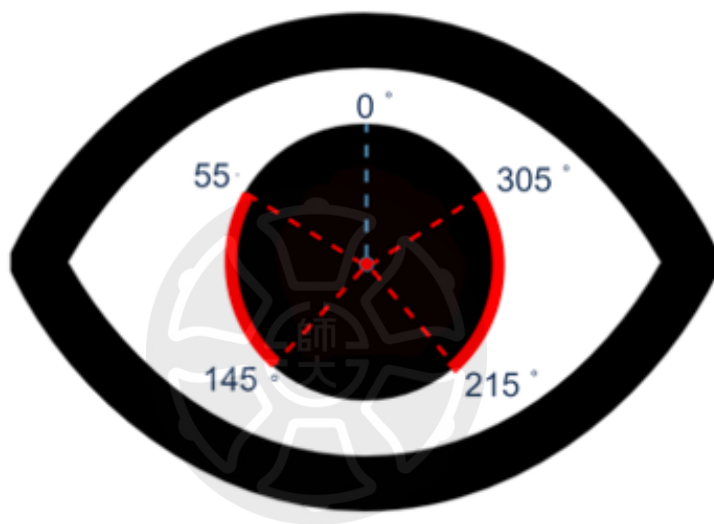


Figure 3.11: 使用者直視相機時所定義的算分區域

直視情境下，也就是眼球旋轉角度 $[\theta_0, \phi_0]$ 估計出來為 $[0, 0]$ 的情況下，將有效算分邊界集中於 55° 至 145° 和 215° 至 305° 的區域，如圖 3.11 所示，分別位於虹膜的左側和右側。這些區域因遠離上下眼皮遮擋而擁有更高的梯度值，因此在算分時具有更高的可信度。

為了能夠因應不同的初始凝視方向動態調整算分範圍，本研究提出了一種基於影像上眼球中心與虹膜中心之間關係的簡單評估方法。此方法能有效地根據投影於影像上的幾何特性，動態決定虹膜邊界的算分區域。

如圖 3.12 所示，該方法通過計算影像上眼球中心 (Eye centroid) 與虹膜中

心 (Iris center) 的水平偏移量 (Horizontal shift) 和垂直偏移量 (Vertical shift)。
Horizontal shift 表示虹膜中心 x 的影像座標值減去眼球中心 x 的影像座標值。
Vertical shift 表示虹膜中心 y 的影像座標值減去眼球中心 y 的影像座標值。並且
 考慮與虹膜初估半徑的對數值之間的比值來實現邊界範圍的調整。以下為水平偏
 移量 R_x 的計算與垂直偏移量 R_y 的計算的公式:

$$R_x = \frac{\text{Horizontal shift}}{\log(\text{iris radius})} \quad (3.11)$$

$$R_y = \frac{\text{Vertical shift}}{\log(\text{iris radius})} \quad (3.12)$$

這些比值的大小以及正負值能夠用來動態確定算分邊界的範圍。例如：當 R_x 為正值時，表示虹膜中心位於眼球中心的右側；當 R_x 為負值時，表示虹膜中心位於眼球中心的左側。當 R_y 為正值時，表示虹膜中心位於眼球中心的下方；當 R_y 為負值時，表示虹膜中心位於眼球中心的上方。

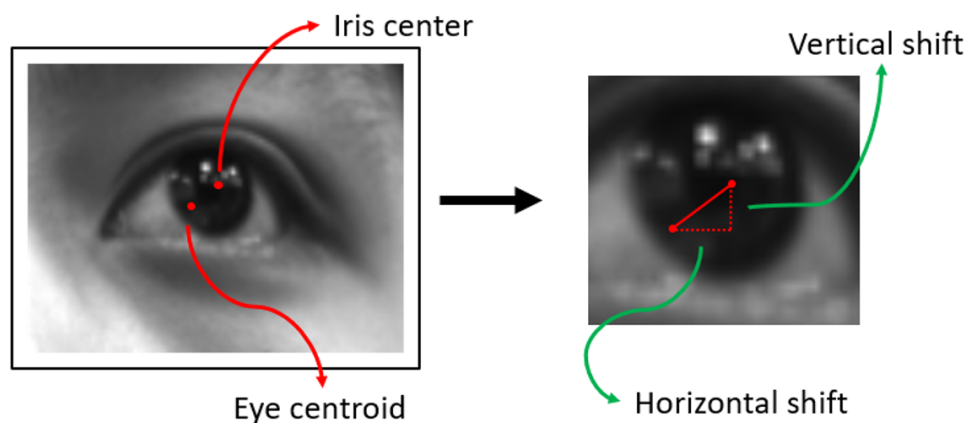


Figure 3.12: 邊界算分方式，根據水平偏移量 R_x 以及垂直偏移量 R_y 與虹膜初估半徑的對數值之間的比值來決定。

基於這些偏移比值的動態判斷，我們定義了一套眼球旋轉的上下左右補償機制。該機制能夠根據虹膜中心與眼球中心的相對關係，自適應地調整算分邊界的

範圍，從而提高視線追蹤系統的穩定性與準確性。如下圖所示：

圖3.13中的 A、B、C、D 分別表示 55° 、 145° 、 215° 和 305° 的位置範圍，作為初始可信邊界。根據圖中表格所示，邊界的調整會隨著虹膜中心與眼球中心之間的相對位置動態變化。

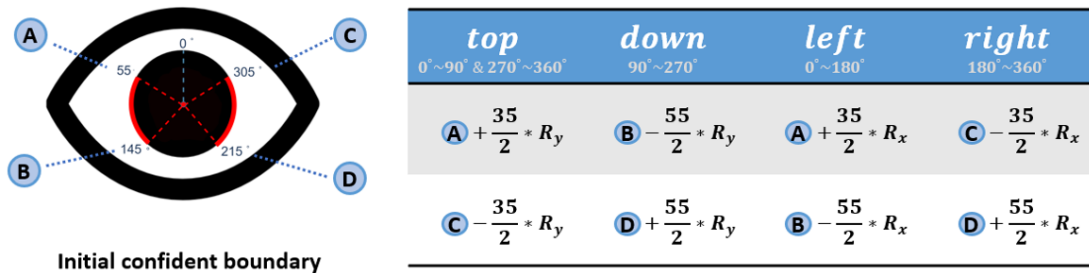


Figure 3.13: 根據虹膜中心與眼球中心的相段關係所設置的補償算分方式，決定當前可信任的邊界。

表格中，依據虹膜中心與眼球中心的水平（Horizontal shift）及垂直（Vertical shift）偏移量比值 R_x 和 R_y ，分別判斷邊界在上下左右四個方向上的變化範圍：Top（上邊界）：範圍在 0° 至 90° 及 270° 至 360° 。Down（下邊界）：範圍在 90° 至 270° 。Left（左邊界）：範圍在 180° 至 360° 。Right（右邊界）：範圍在 0° 至 180° 。根據表格中的數學表達式，每個邊界的範圍會因虹膜中心的偏移量而產生動態調整。例如當 R_y 增大時，上或下邊界可能發生變動。當 R_x 增大時，左右邊界會根據比值變化進行相應調整。

這種設計能動態決定各方向邊界的可信範圍，從而提升系統在不同凝視方向下的穩定性與準確性。

3.4.3 虹膜匹配

我們會根據子章節 3.4.1 中求得的 θ, ϕ 範圍，作為搜索的初始範圍，進行最佳眼球旋轉角度的計算。在該範圍內，我們會計算每一組 θ, ϕ 所對應的 Fitness Score。Fitness Score 的定義如下：

$$f(\theta, \phi) = \prod_{\alpha \in \mathcal{B}} \Delta p_{\theta, \phi}(\alpha) \quad (3.13)$$

其中， $\Delta p_{\theta,\phi}(\alpha)$ 表示在角度 α 射線方向上，影像像素梯度變化的強度， B 是上一章節所計算出的虹膜邊界範圍的角度集合。基於子章節 3.4.1 所縮小的 θ, ϕ 範圍，我們能夠有效地排除不相關的干擾區域，專注於對應視線方向的關鍵邊界進行精細計算。

接著，我們在這個範圍內對每一組 θ, ϕ 組合計算 Fitness Score，並選取分數最高的組合作為最佳匹配點。這對應於虹膜邊界與模型的最佳適配位置，即我們所要求的最佳眼球旋轉角度。透過這一方法，我們不僅能夠顯著提升視線估計的精確性，還減少了因大範圍搜索導致的運算資源消耗。

此外，結合子章節 3.4.1 中的邊界範圍選擇策略，我們的 Fitness Score 計算方法進一步優化。例如，對於梯度變化值 $\Delta p_{\theta,\phi}(\alpha)$ ，當其小於某一閾值時，會設定為零以過濾噪聲干擾，確保計算的穩定性和準確性。這樣的方法既提高了模型的計算效率，又為多種動態條件下的視線追蹤應用奠定了基礎。

3.5 視線方向與螢幕位置的映射關係

在視線追蹤技術中，螢幕與視線的映射模型對於準確性和穩定性至關重要。常用的映射方法包括多項式回歸和偵錯點方法，這兩者在應對不同場景和數據特性時各有優勢。多項式回歸是一種非線性回歸方法，能夠捕捉視線與螢幕座標之間的複雜非線性關係，因此在使用者的視線分布存在非線性特性時，該方法的準確性更高。然而，多項式回歸的缺點在於對異常點的敏感性。如果數據中存在偏差較大的點（例如使用者未準確注視校正點），模型的整體效果可能會受到嚴重影響。

與此相比，偵錯點方法是一種線性映射方法，其核心精神在於利用少量穩定的視點快速建立初步公式，並通過迭代排除數據中的異常點以逐步提高模型的準確性。具體而言，該方法首先從九個校正點中選擇三個偏差最小且分佈均勻的視點作為基礎，建立初步的線性回歸公式。接下來，利用該公式計算其他校正點的

偏差，將偏差過大的點標記為異常點並排除，然後重新選擇穩定點進行公式迭代。這樣的迴圈過程能逐步剔除不準確的數據點，最終形成一個穩定且高效的映射公式。偵錯點方法的優勢在於其對異常點的容忍度較高，適合數據中包含偏差點或異常點的情況，但在視線分布呈現顯著非線性時，可能不如多項式回歸精確。

為了決定在不同場景中選擇哪種映射方法，可以基於均方差（Mean Squared Error, MSE）來評估兩種方法的表現。具體步驟包括，首先對所有九個校正點的數據使用多項式回歸，計算其均方差（ MSE_{poly} ）；其次，採用傾斜點方法，選取三個穩定點建立初步公式，排除異常點後重新計算最終公式的均方差（ MSE_{tilt} ）。最後，通過比較兩種方法的均方差來選擇最合適的方法。如果 MSE_{poly} 小於 MSE_{tilt} ，則選擇多項式回歸，因為數據特性更適合非線性映射；反之，若 MSE_{poly} 大於 MSE_{tilt} ，則選擇傾斜點方法，因為該方法能更好地處理異常點並提高整體模型穩定性，下圖3.14為具體的流程圖。

總體來說，多項式回歸與傾斜點方法各有其適用場景。多項式回歸在視線分布具有顯著非線性特性時表現優越，但對異常點敏感；傾斜點方法則以簡單高效和容錯性高為特點，特別適合於數據中包含偏差點的情況。通過結合兩種方法的特性，並利用均方差作為選擇標準，可以靈活應對不同的視線數據特性，最終實現更準確、更穩定的視線映射。

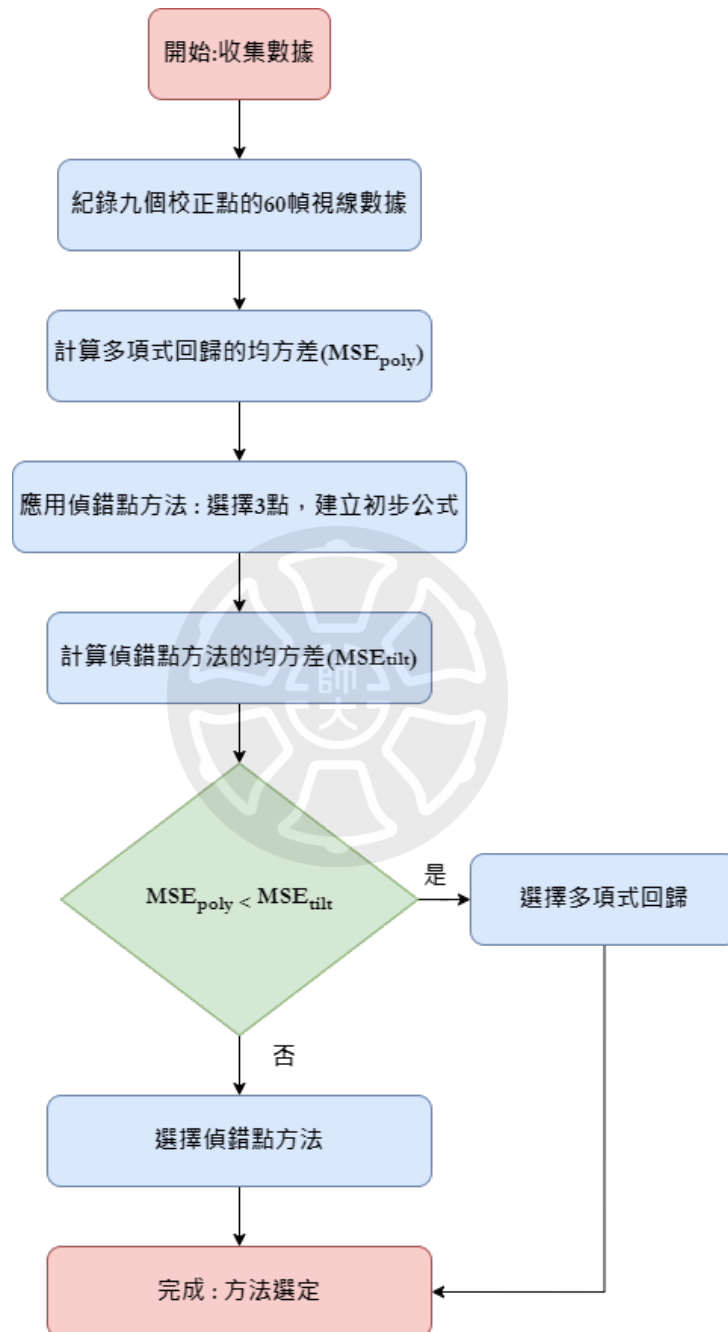


Figure 3.14: 映射公式判斷流程圖

第四章 實驗結果

4.1 實驗環境與系統設置

本研究採用 XIMEA 品牌的高速攝像機（型號：MQ042MG-CM）作為核心影像捕捉設備，如圖4.1，其影像解析度達 2048 x 2048 像素，能以每秒 90 幀的速度進行影像擷取。實驗顯示設備則選用 27 吋顯示器，並使用 AMD Ryzen 7 3800X 作為實驗平台的中央處理器，提供穩定且高效的計算能力。



Figure 4.1: XIMEA MQ042MG-CM

實驗在一間普通的室內辦公室環境中進行，為保持自然的測試條件，未特意排除室內燈光影響，以模擬真實使用場景下的環境光源分佈。相機位置設置於顯示器下方，並配備主動對焦功能，確保能覆蓋使用者的整張臉部。相機的焦距經過調整，能適應使用者自然、舒適觀看螢幕的範圍。

為了保證實驗數據的準確性，實驗過程中會確認相機能完整捕捉使用者面部的全貌並始終保持精確對焦。此外，影像捕捉的設定經過細緻的調校，以確保眼球運動的關鍵特徵點能被清晰記錄，從而為後續的分析提供可靠數據。

整體而言，實驗設備的設置旨在平衡功能性與使用者舒適性，同時模擬真實的使用環境，為研究提供了穩定且高效的測試條件，下圖4.2為實驗環境示意圖。



Figure 4.2: 環境配置圖

4.2 頭部姿態估計實驗

為了驗證 Levenberg-Marquardt (LM) 演算法在 solvePnP 問題中的優勢，我們設計了一組實驗，對比 LM、梯度下降法 (GD)、牛頓法 (Newton) 和高斯-牛頓法 (GN) 的性能表現。這些比較主要圍繞三個核心指標進行：收斂時間、精度以及穩定性。在實驗設計中，我們模擬了兩種不同初始條件場景，即良好初始條件和差初始條件，並評估各算法在這兩種情況下的表現。

以下表格展示了四種方法在不同條件下的具體實驗結果：

首先，在收斂時間方面，實驗結果顯示，LM 和 GN 在良好初始條件下表現出最快的收斂速度，分別為 15.3 ms 和 16.8 ms。這說明這兩種方法能夠在接近真實解的情況下有效地完成最小化優化。然而，當初始條件不佳時，LM 的收斂時間為 28.5 ms，顯著優於 GN 的 35.2 ms，並且遠超梯度下降法和牛頓法的表現。GD 在此情況下未能收斂，而牛頓法則因初始猜測遠離全局解導致數值不穩定，無法有效完成優化。

在精度方面，LM 方法在兩種初始條件下的 RMSE 均為最低，分別為 0.12 和 0.15，顯示其強大的誤差最小化能力。GN 方法的 RMSE 次之，而牛頓法在良好

Table 4.1: 四種數值演算法在 solvePnP 問題中的性能比較

方法	初始條件	收斂時間 (ms)	RMSE (像素)	穩定性 (% 成功收斂)
LM	良好	15.3 ± 1.2	0.12 ± 0.03	98%
	差	28.5 ± 2.8	0.15 ± 0.05	92%
GD	良好	45.7 ± 5.4	0.18 ± 0.07	72%
	差	不收斂	-	0%
Newton	良好	18.9 ± 2.3	0.14 ± 0.04	90%
	差	不穩定	-	40%
GN	良好	16.8 ± 1.9	0.13 ± 0.05	94%
	差	35.2 ± 3.6	0.20 ± 0.08	70%

條件下能達到接近 GN 的精度，但在差初始條件下卻表現不穩定。梯度下降法的精度最差，並且在差初始條件下無法提供有效解。這表明，LM 不僅在速度上具有優勢，還能更精確地求解 solvePnP 問題。

最後，從穩定性角度分析，LM 的優勢尤為明顯。在良好初始條件下，LM 和 GN 均有超過 90% 的成功收斂率，而梯度下降法和牛頓法分別只有 72% 和 90%。在差初始條件下，LM 的穩定性依然保持在 92%，而 GN 則降至 70%，牛頓法僅有 40% 的成功率，GD 則完全無法收斂。這表明，LM 在初始猜測較差的情況下，能夠通過其混合的數值更新機制，動態平衡步長與方向，從而保持高穩定性。

綜上所述，LM 演算法在解決 solvePnP 問題時展示出顯著的性能優勢。它能夠在收斂速度和最小化精度之間實現良好的平衡，並且在初始條件不佳時提供顯著更高的穩定性。與其他數值方法相比，LM 的表現證明了其在處理此類非線性

最小化問題中的卓越能力。因此，基於這些實驗結果，我們選擇 LM 作為本研究中 solvePnP 問題的數值解法。

4.3 虹膜分割模型實驗與結果

本研究基於 RITnet [5] 模型框架進行改進，該框架結合了 DenseNet 和 U-Net 的優勢，具有輕量化和高效性能的特點。以下將詳細介紹實驗設計、數據集、模型配置及實驗結果。

4.3.1 數據集處理與修改

本研究採用 UBIRIS.v2[17] 數據集作為模型的訓練、驗證與測試數據來源。該數據集原始的標註（ground truth）僅對虹膜部分進行標記，瞳孔被視為背景。然而，考慮到亞洲人眼部特徵中，瞳孔與虹膜的顏色通常較為相近，這樣的標註可能導致模型在分割邊界時產生混淆，進而影響模型性能。因此，本研究對 UBIRIS.v2 的原始標註進行了調整，將瞳孔與虹膜視為同一區域進行標記。

圖 4.3 示範了修改過程及其效果。上排為原始標註結果，僅包含虹膜區域，瞳孔部分被標記為背景；下排則為修改後的標註結果，將瞳孔與虹膜合併為同一區域。這樣的處理有助於模型更準確地學習虹膜整體的特徵，並提高分割的準確性和邊界的一致性。

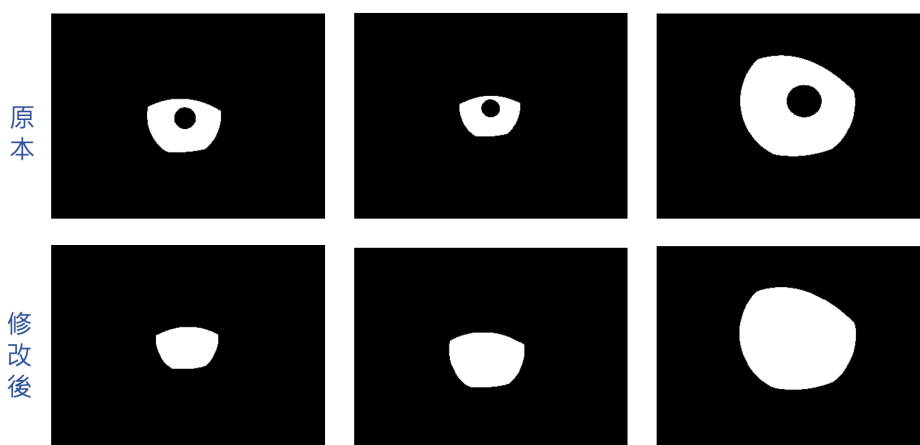


Figure 4.3: 數據集 ground truth 修改示例。上排為原始標註，僅包含虹膜；下排為修改後標註，將瞳孔與虹膜視為一致。

4.3.2 模型預處理和訓練

在數據預處理階段，我們採用了 gamma 校正和 CLAHE（對比度受限的自適應直方圖均衡化）技術，以提升圖像品質並強化虹膜邊界特徵。為了進一步模擬不同環境條件，數據增強技術包括隨機旋轉、水平翻轉、高斯模糊和隨機遮擋。

改進的 RITnet 模型以輕量化為目標，總參數數量僅 248,900，模型大小不足 1 MB，適合資源有限的應用場景。模型採用了改進的損失函數設計，結合交叉熵損失（Cross Entropy Loss, CEL）與邊界感知損失（Boundary Aware Loss, BAL），提高了分割結果的準確性和邊界清晰度。訓練過程中，優化器使用 Adam，初始學習率為 0.001，並在驗證集損失無改進時逐步衰減學習率。訓練批量大小設為 16，共進行 68 個 epoch。在 NVIDIA RTX 3070 Ti GPU 上，每張影像的推論時間為 10 毫秒，能實現每秒超過 100 幀的實時處理能力。

4.3.3 量化評估與結果

在測試集上的性能評估中，我們使用了準確率（Accuracy）和 IoU（Intersection over Union）作為主要指標。以下對兩者進行說明：

準確率（Accuracy）是衡量模型整體預測正確性的指標，計算公式如下：

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (4.1)$$

其中，TP 為正確預測的正類像素數，TN 為正確預測的負類像素數，FP 和 FN 分別為錯誤預測的正類和負類像素數。準確率反映了整體像素的正確率，適合評估模型在所有像素上的表現。

IoU（Intersection over Union）用於衡量模型預測區域和真實區域的重疊程度，計算公式如下：

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (4.2)$$

IoU 更關注模型對目標區域（如虹膜）的分割精度，能有效反映模型的局部性

能。

以下表格展示了模型在訓練集、驗證集和測試集上的表現：

Table 4.2: 改進的 RITnet 模型在各數據集上的表現

指標	訓練集	驗證集	測試集
準確率 (Accuracy)	98.7%	96.8%	95.2%
IoU	92.5%	90.3%	89.1%

此外，我們將模型的分割結果與輸入影像進行了可視化對比，如圖 4.4 所示。從圖可以看出 RITnet 模型能夠精確分割虹膜區域，並保持邊界的完整性與清晰度。

從實驗結果可以看出，改進的 RITnet 模型在測試集上的準確率達到 95.2%，IoU 為 89.1%，充分展示了其優異的分割性能。此外，損失函數的改進有效提升了分割邊界的準確性，訓練過程中損失曲線和準確率曲線表明模型在第 50 個 epoch 後趨於穩定，並在第 68 個 epoch 達到最佳性能。

總結來說，該模型的輕量化設計使其在實時應用場景中具有較大的優勢，例如視線追蹤和身份識別等。該研究提出的改進方法不僅提升了虹膜分割的準確性，還為多樣化場景中的應用提供了穩健支持。

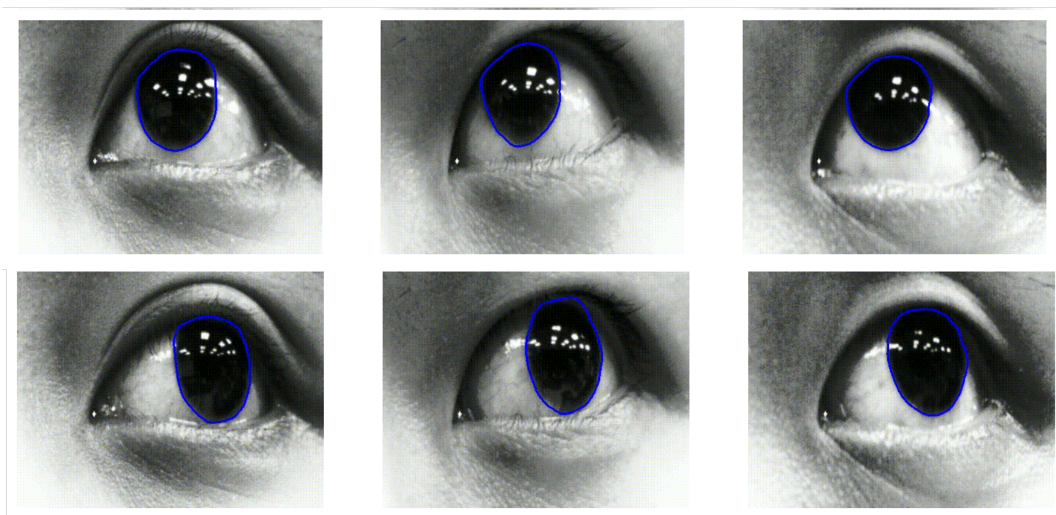


Figure 4.4: 測試資料虹膜分割結果圖

4.4 虹膜匹配優化實驗

本實驗旨在驗證“搜尋範圍優化”和“算分區域”兩項改進對於視線追蹤系統性能的提升效果，分別從計算效率和匹配精度兩個角度進行量化分析。

4.4.1 階層搜尋方法

階層搜尋分為三個階段，每階段逐步縮小搜尋範圍並減小步進值，以降低計算負擔並提升精度。各階段的參數設置如表 4.3 所示：

Table 4.3: 階層搜尋方法參數設置

Phase	$\Delta\alpha$	$\Delta\theta$	$\Delta\phi$	Search range	Choose the top
1	18	1.5	5	$\theta \in [0, \frac{\pi}{2}], \phi \in [0, \frac{\pi}{2}]$	10
2	12	1	2.5	$\dot{\theta} \in (\theta_r - 3, \theta_r + 3), \dot{\phi} \in (\phi_r - 5, \phi_r + 5)$	5
3	9	0.5	0.5	$\dot{\theta} \in (\theta_r - 1, \theta_r + 1), \dot{\phi} \in (\phi_r - 2, \phi_r + 2)$	1

在第一階段，我們對 θ 和 ϕ 進行大範圍搜尋，步進值較大 ($\Delta\theta = 1.5^\circ, \Delta\phi = 5^\circ$)，並選取 Fitness Score 前 10 名進入下一階段；第二階段進一步縮小範圍，步進值減小至 $\Delta\theta = 1^\circ, \Delta\phi = 2.5^\circ$ ，選取前 5 名候選結果；最終階段中，搜尋範圍被進一步縮小，步進值減至 0.5° ，並選取得分最高的結果作為最佳匹配點。

4.4.2 搜尋範圍優化對計算效率的影響

本實驗比較了傳統方法與搜尋範圍優化方法的計算時間。兩者均採用階層搜尋方法，但傳統方法的初始搜尋範圍為 $\theta \in [0^\circ, 30^\circ]$ 和 $\phi \in [0^\circ, 180^\circ]$ ，而搜尋範圍優化的方法則將初始範圍縮小至 $\theta \in [\theta_{\text{initial}} - 5^\circ, \theta_{\text{initial}} + 5^\circ]$ 和 $\phi \in [\phi_{\text{initial}} - 10^\circ, \phi_{\text{initial}} + 10^\circ]$ 。表 4.4 展示了兩種方法的平均計算時間對比。

數據顯示，搜尋範圍優化將平均計算時間從 0.02427 秒降低至 0.01836 秒，減

Table 4.4: 搜尋範圍優化對計算時間的影響

方法	初始搜尋範圍	平均計算時間 (秒)
傳統方法	$[0^\circ, 30^\circ], [0^\circ, 180^\circ]$	0.02427
搜尋範圍優化	$[\theta_{\text{initial}} \pm 5^\circ], [\phi_{\text{initial}} \pm 10^\circ]$	0.01836

少了約 24.3%。這表明縮小搜尋範圍在減少計算負擔方面具有顯著效果。

4.5 凝視估計系統性能評估

為了評估凝視估計系統的性能，通常會採用兩個指標：精準度 (Accuracy) 和精密度 (Precision)。這兩個指標基於視角轉換 (Visual angle)，因為視角是一個能夠統一量化不同距離測量結果的標準化參數。由於凝視方向的偏差可能受使用者頭部位置改變或環境因素影響，因此基於視角的評估方法能有效比較不同距離或不同條件下系統的性能。

4.5.1 像素視角轉換

基於凝視估計的需求，我們需將螢幕上的像素距離轉換為視角距離 (Pixel to Visual Angle, P2A)。這種轉換提供了統一的基準來量化凝視偏差，公式如下：

$$P2A(d_{\text{pixel}}) = 2 \times \arctan \left(\frac{d_{\text{pixel}}}{2 \cdot d_{\text{user-to-monitor}}} \right) \quad (4.3)$$

其中， d_{pixel} 表示像素距離， $d_{\text{user-to-monitor}}$ 為用戶眼睛與螢幕的距離。

4.5.2 系統精準度

凝視估計系統的精準度衡量系統計算的凝視點與目標點之間的平均視角誤差，其公式如下：

$$\text{Accuracy} = P2A \left(\frac{\sum_{i=1}^N \sum_{j=1}^M |G_{ij} - R_{ij}|}{N \times M} \right) \quad (4.4)$$

其中， G_{ij} 為目標點的位置， R_{ij} 為系統估算得到的凝視點位置， N 是目標點數量， M 是每個目標點的測試次數。

4.5.3 系統精密度

凝視估計系統的精密度則是量化系統輸出凝視點穩定性的指標，其計算公式為：

$$\text{Precision} = \sqrt{\frac{\sum_j^K (P2A(R_{ij}) - \bar{R})^2}{K}} \quad (4.5)$$

其中， $P2A(R_{ij})$ 表示測得的凝視點轉換為視角的結果， \bar{R} 為凝視點的平均值， K 為測試次數。

4.6 熱區圖結果對比

本章節基於使用者觀看螢幕九點後的凝視點分布，透過熱區圖（heatmap）進行結果對比分析，評估本研究提出的方法與舊有方法的性能差異。熱區圖提供了一種直觀的方式，用於呈現使用者目光的集中區域與分布特性，能有效展示系統穩定性與精確性。本研究的核心目標是驗證提出的方法是否能在凝視點分布上表現出更高的一致性與準確度。以下將詳細說明熱區圖結果的量化與解讀。

4.6.1 九點驗證

本研究提出的方法以改進先前研究的方法為基礎，解決視線追蹤中因頭部運動而導致精度下降的問題。為了展示新舊方法的差異，我們利用熱區圖進行比較分析。熱區圖呈現了使用者在進行視線追蹤任務時，觀看螢幕上九個參考點時的凝視點分布。我們收集了 15 位使用者的數據作為分析基礎。實驗設計包括校準階段和驗證階段，以確保結果的精確性與可靠性。

在校準階段，使用者被要求盡量保持頭部穩定，依序直視相機以及觀看螢幕上的九個點，以建立準確的凝視點映射公式。其中，直視相機時會記錄 30 幀影像，而每個觀看九個點的階段則各記錄 60 幀影像。完成校準並建立映射公式後，

進入驗證階段。此階段允許使用者輕微轉動頭部，以模擬更自然的觀看九點的行為，從而檢驗方法的實用性與適應性。在驗證階段，九個點各記錄 60 幀數據，作為結果分析的依據。下圖為驗證階段觀看九點的示意圖，可以發現使用者會自然地調整頭部至觀看九點的位置。

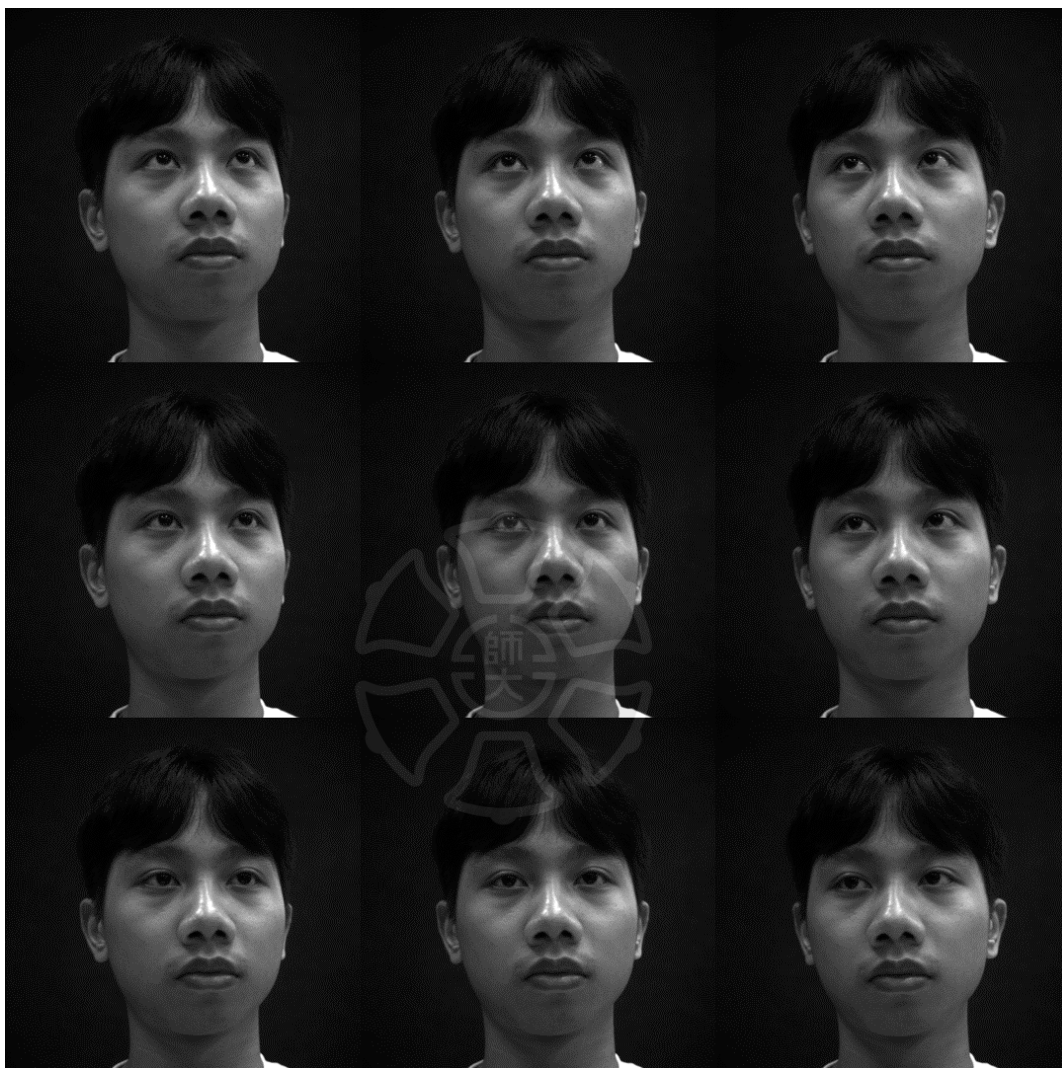
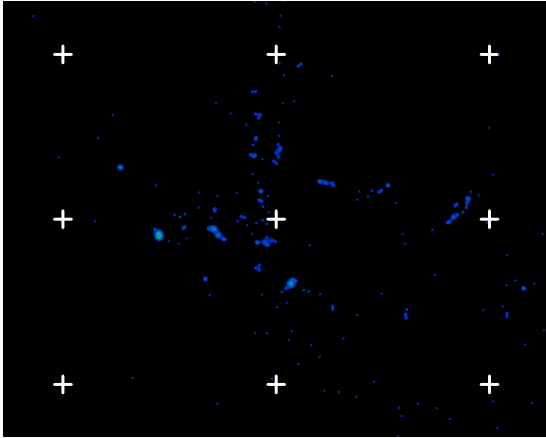
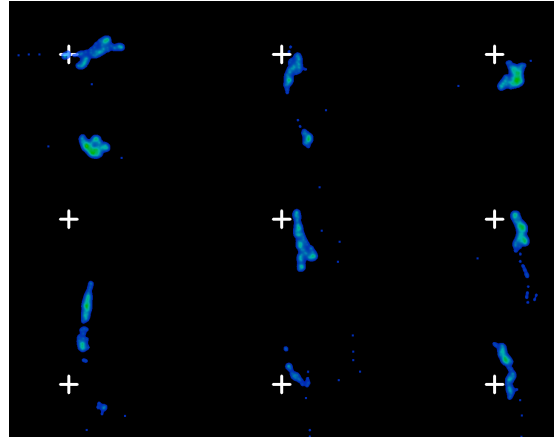


Figure 4.5: 驗證階段觀看九點的使用者狀態，

接著我們對其中幾筆 data 做分析。首先，針對 data2 和 data3 的分析結果，可以觀察到舊有方法因系統不穩定性，導致校準階段無法建立穩定且可靠的映射公式，從而在驗證階段的精密度與精確度表現均受到限制，如圖4.6和圖4.7。然而，採用本研究提出的新方法後，即使在相同的條件下，仍能成功建立穩定且準確的映射公式，並在驗證階段呈現出明顯改善的結果，充分展現出方法的可靠性與穩定性。



(a) 補償前的熱區圖

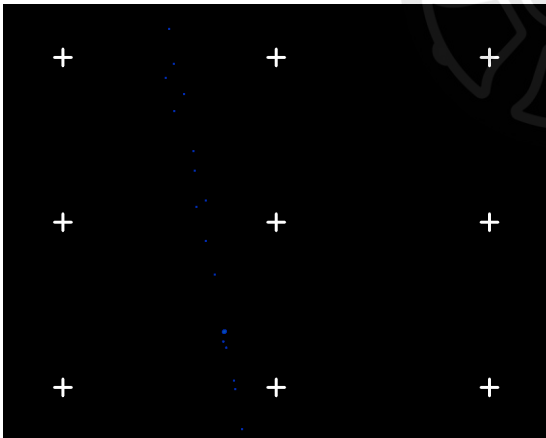


(b) 補償後的熱區圖

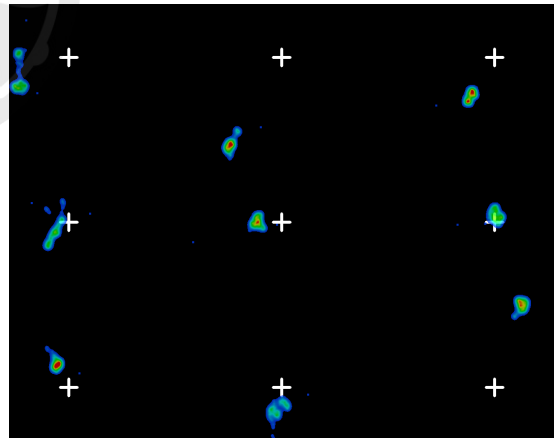
Figure 4.6: data2 : 前後的熱區圖結果對比。

Table 4.5: Data 2 精確度與精準度比較

Data 2	Baseline method	Proposed method
Accuracy	55.6	2.14
Precision	46.46	1.03



(a) 補償前的熱區圖



(b) 補償後的熱區圖

Figure 4.7: data3 : 前後的熱區圖結果對比。

Table 4.6: Data 3 精確度與精準度比較

Data 3	Baseline method	Proposed method
Accuracy	56.6	2.25
Precision	47.46	0.54

接下來，針對 data6 和 data7 的分析顯示，這兩組數據在校準階段均成功建立了高精度的映射公式。在驗證階段中，應用了頭部姿態估計補償的新方法相比舊方法，精密度雖然相似，但精確度有顯著提升，表明新方法能有效補償頭部轉動對凝視點分布的影響，如圖4.8和圖4.9。這進一步驗證了本研究方法在適應使用者自然頭部運動方面的優勢。

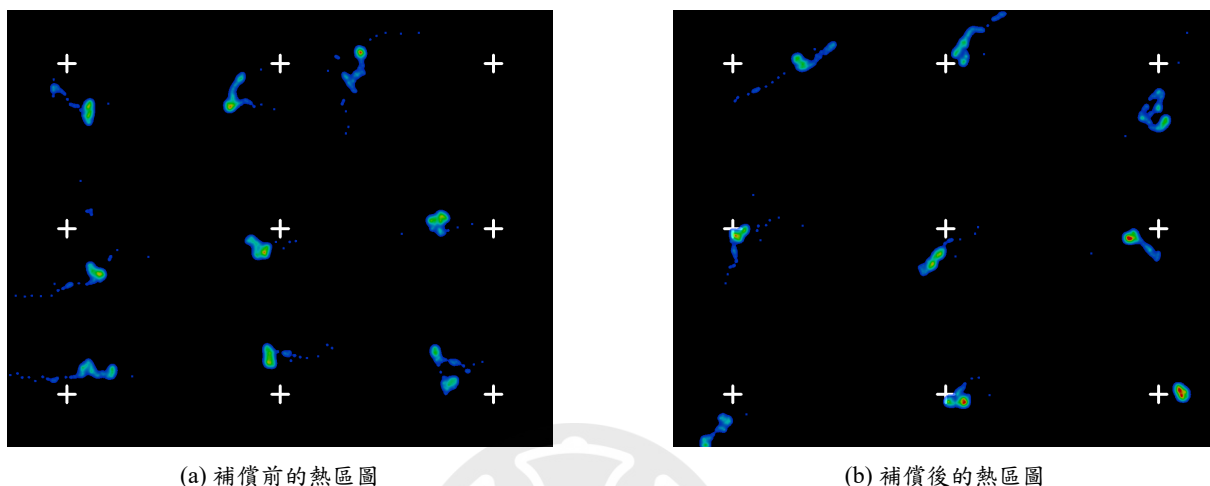


Figure 4.8: data6 : 前後的熱區圖結果對比。

Table 4.7: Data 6 精確度與精準度比較

Data 6	Baseline method	Proposed method
Accuracy	2.87	1.80
Precision	0.65	0.67

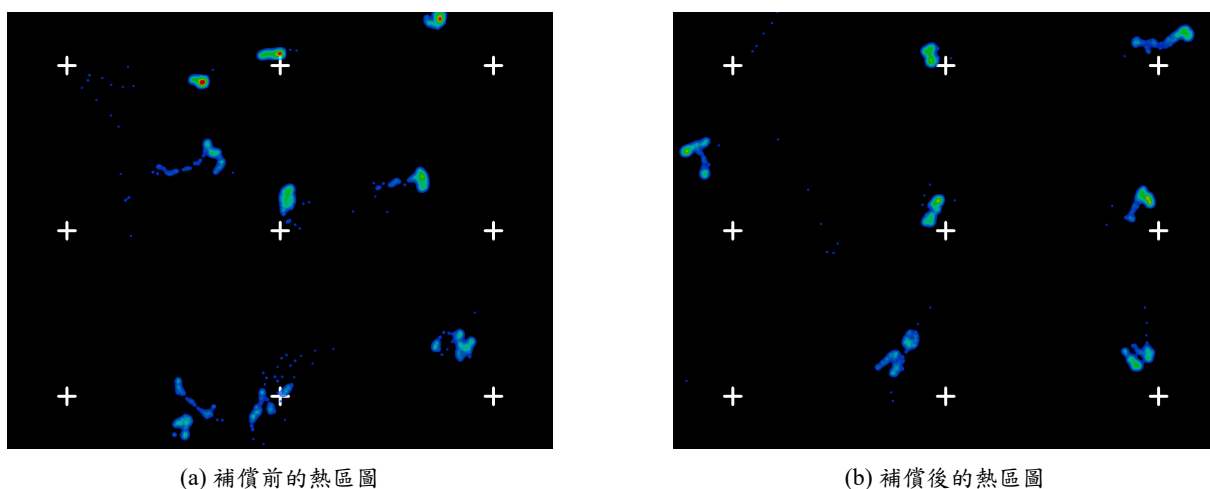


Figure 4.9: data7 : 前後的熱區圖結果對比。

Table 4.8: Data 7 精確度與精準度比較

Data 7	Baseline method	Proposed method
Accuracy	4.03	3.03
Precision	0.72	0.69

4.6.2 十七點驗證

在進行使用者凝視評估時，除了以九點作為基準，我們額外設計了十七個凝視點來進行驗證。最初，使用者需保持頭部穩定並注視螢幕以完成校準。隨後，在驗證階段，使用者會被要求旋轉頭部一定角度後，再次注視這十七個驗證點。如下圖4.10。

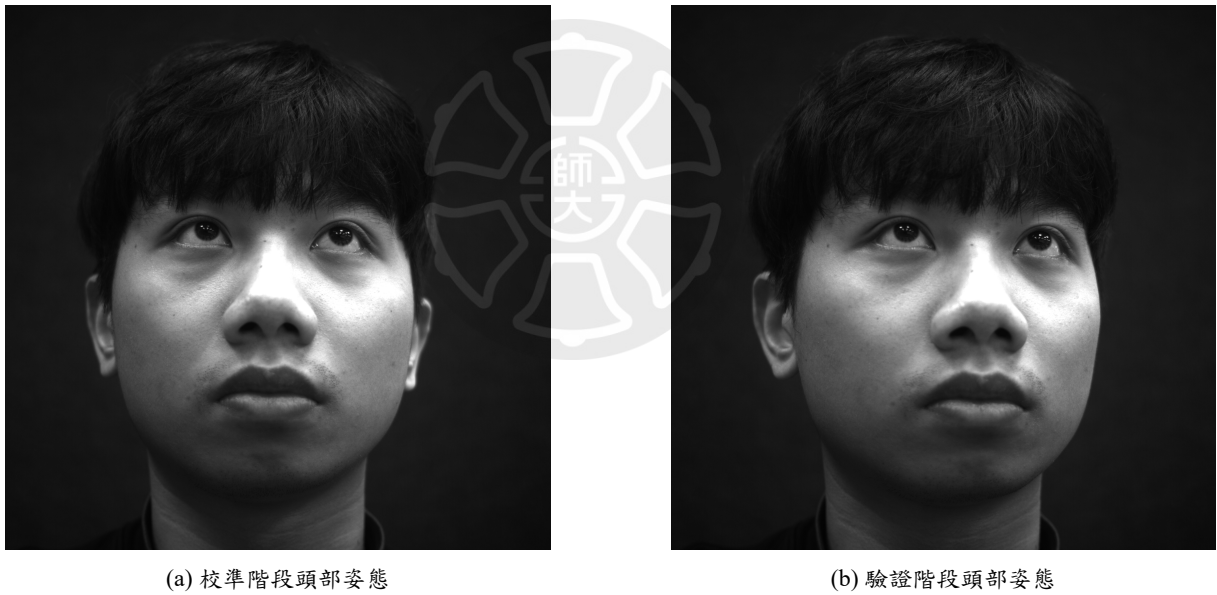
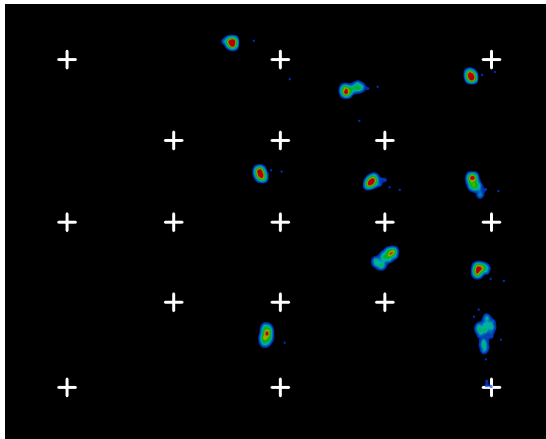
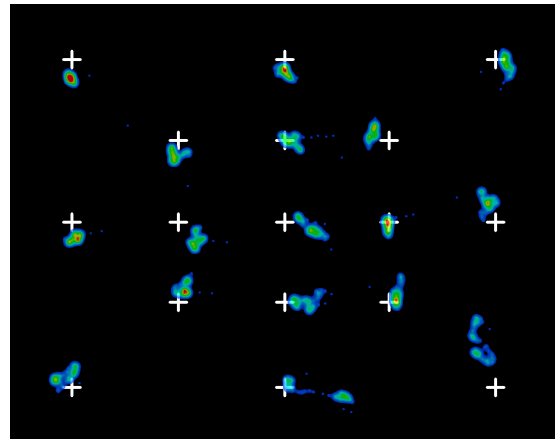


Figure 4.10: 。

從 data19 和 data28 的驗證結果來看，如圖4.11及圖4.12，傳統的 baseline 方法表現不佳，計算出的凝視點往往偏離正確位置，甚至超出視窗範圍。然而，我們提出的方法能夠更有效地抑制因頭部旋轉所導致的偏差，使預測的凝視點更接近真實目標點，大幅提升準確性與穩定性。



(a) baseline : 17 點驗證熱區圖

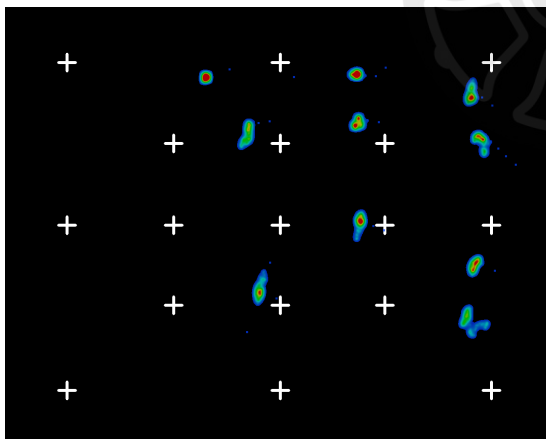


(b) Proposed method : 17 點驗證熱區圖

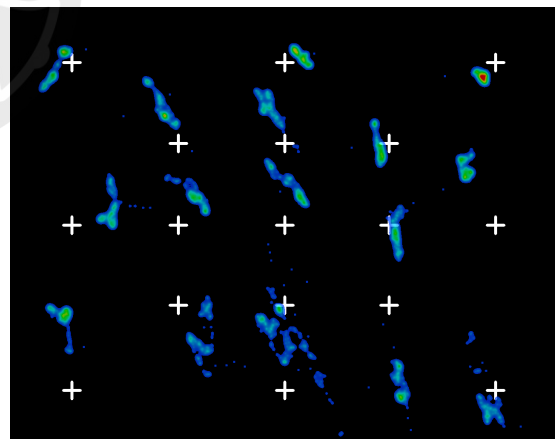
Figure 4.11: data19 : 熱區圖比較

Table 4.9: Data19 驗證階段精確度與精準度比較

Data 19	Baseline method	Proposed method
Accuracy	30.90	0.94
Precision	22.31	1.30



(a) baseline : 17 點驗證熱區圖



(b) Proposed method : 17 點驗證熱區圖

Figure 4.12: data28 : 熱區圖比較

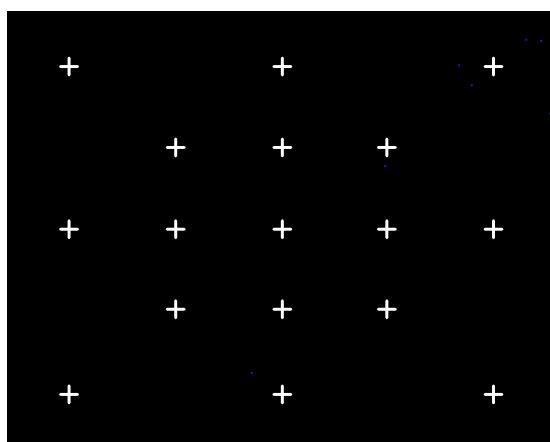
Table 4.10: Data28 驗證階段精確度與精準度比較

Data 28	Baseline method	Proposed method
Accuracy	31.20	1.94
Precision	22.47	1.55

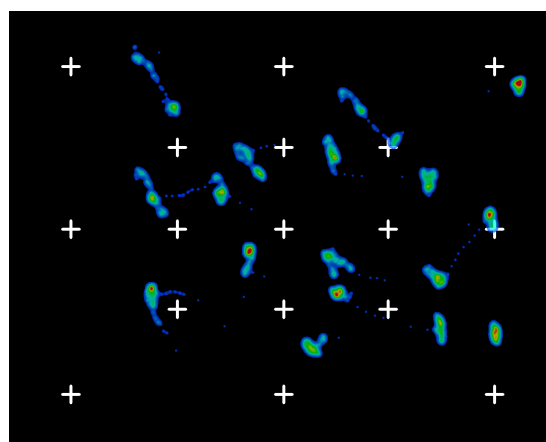
由於不同使用者的眼部構造和視線習慣存在個體差異，視線追蹤數據可能會產生一定偏差，特別是在頭部旋轉時，這種偏差可能會進一步放大，導致視線落點與預期不符。例如，在 data24 和 data29 的數據中，雖然 17 個凝視點都位於背景圖上，但其分布並未達到理想狀態，顯示出視線偏移可能影響結果的準確性。如圖 4.13 和圖 4.14。因此，在處理視線追蹤數據時，需要考慮這些變因並進行補償或校正，以確保數據的一致性與可靠性。

提出的補償方法能有效修正因視線偏差與頭部旋轉所產生的誤差，使視線落點更接近理想分布。然而，仍有部分使用者即使存在頭部角度變化，卻能自然地達到理想狀態，這可能與個體視線穩定性、頭部與眼睛的協調性或視線追蹤設備的適應性有關。例如，一些使用者的眼動控制較為精確，或在頭部旋轉時會自動調整眼球運動，減少誤差，使其數據看起來較為準確。

因此，這些個體的數據雖然未經補償仍能達到理想狀態，但這並不代表補償方法無效。所提出的方法能夠有效改善那些受影響較大的使用者數據，使整體視線落點分布更為準確，提升視線追蹤結果的穩定性與可預測性，確保應用於後續分析時的可靠性。



(a) baseline : 17 點驗證熱區圖

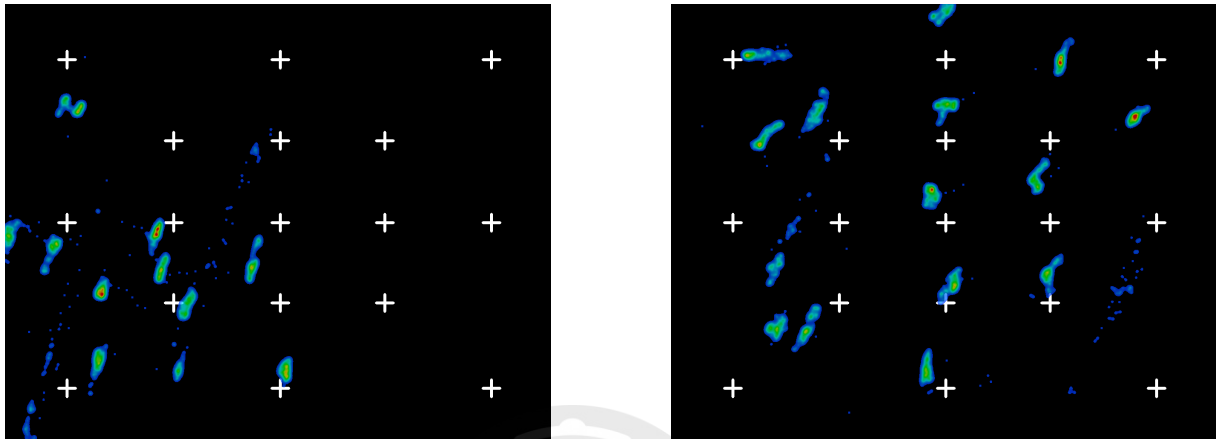


(b) Proposed method : 17 點驗證熱區圖

Figure 4.13: data24 : 熱區圖比較

Table 4.11: Data24 驗證階段精確度與精準度比較

Data 24	Baseline method	Proposed method
Accuracy	44.47	3.40
Precision	38.31	1.21



(a) baseline : 17 點驗證熱區圖

(b) Proposed method : 17 點驗證熱區圖

Figure 4.14: data29 : 熱區圖比較

Table 4.12: Data29 驗證階段精確度與精準度比較

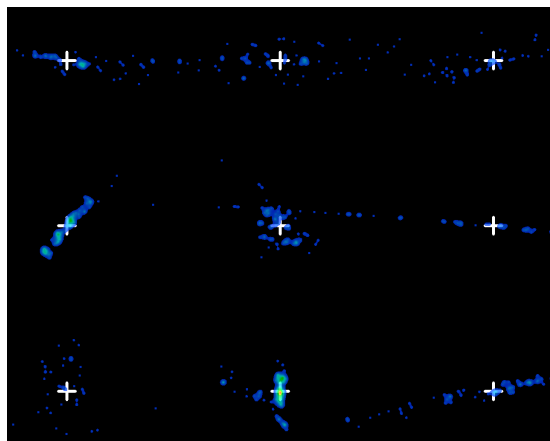
Data 29	Baseline method	Proposed method
Accuracy	71.33	23.43
Precision	64.44	22.61

4.6.3 映射方法在校準的差異

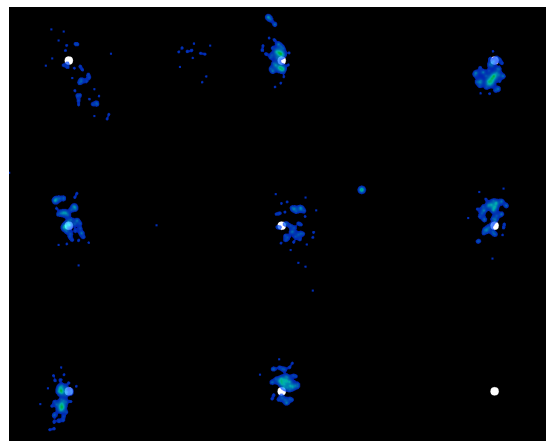
另外在校準階段針對 data5 以及 data10 的分析中，我們觀察到部分凝視點存在校準不準的情況，這導致直接採用多項式回歸生成的映射公式在均方誤差 (MSE_{poly}) 上表現不佳。通過映射方法的判斷流程進行選擇，計算偵錯點方法生成的線性映射公式的均方誤差 (MSE_{edit}) 後發現， MSE_{edit} 顯著低於 MSE_{poly} ，因此最終選擇了偵錯點方法作為映射方式。

驗證階段的結果顯示，採用偵錯點方法後，映射公式的穩定性與精度均顯著

提升，修正了校準階段中凝視不準點所帶來的偏差。相較於多項式回歸方法，偵錯點方法能更有效處理數據中局部偏差的影響，驗證階段的數據誤差縮減幅度達到了明顯優化，如圖4.15 和圖4.16。



(a) 多項式回歸

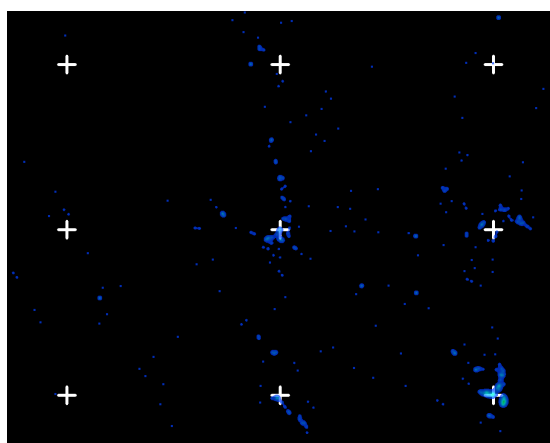


(b) 偵錯點法的線性映射

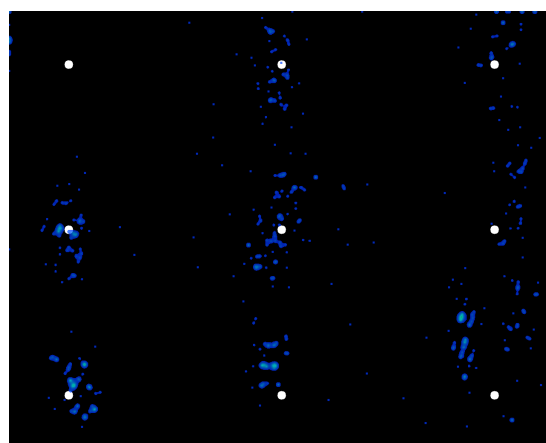
Figure 4.15: data5 : 校準階段不同映射方式的差異。

Table 4.13: Data5 驗證階段精確度與精準度比較

Data 5	多項式回歸	偵錯點法的線性映射
Accuracy	4.03	1.2
Precision	0.72	1.02



(a) 多項式回歸



(b) 偵錯點法的線性映射

Figure 4.16: data5 : 校準階段不同映射方式的差異。

Table 4.14: Data 10 驗證階段精確度與精準度比較

Data 10	多項式回歸	偵錯點法的線性映射
Accuracy	54.13	8.12
Precision	54.13	3.1

這些結果說明，在校準階段數據品質受影響的情況下，基於均方誤差的判斷流程能夠有效選擇出更適合的映射方法，從而保證驗證階段的準確性與穩定性。

綜合上述分析結果，本研究提出的方法不僅在精密度與精確度上均有所提升，還能有效應對使用者頭部運動帶來的挑戰。無論是在校準階段的穩定性，還是驗證階段的表現，新方法均展現出其在視線追蹤應用中的卓越潛力。



第五章 結論與未來展望

5.1 結論

本研究證明，透過結合頭部姿態估計與精確的虹膜分割技術，能有效提升視線追蹤系統在真實世界中的準確性與穩健性。特別是在使用者進行頭部移動的情況下，所提出的方法展現出優異的適應能力。該系統僅需單一相機即可運行，具備高效能表現且具成本效益，為實用性的視線追蹤技術提供了一種可行的解決方案。相比傳統複雜的系統，這種方法簡化了硬體設置，並在降低成本的同時，顯著提升了應用靈活性，成為一種強大的替代方案。本研究的結果顯示，此技術在動態環境下的應用潛力十分廣泛，特別是在需要高準確度和穩定性的情境中具有實際價值。

5.2 未來展望

本研究展現出視線追蹤技術在真實世界應用中的巨大潛力，未來我們將從多個方向進一步優化與拓展，推動此技術邁向更廣泛的實際應用。首先，我們將致力於提升頭部姿態估計與虹膜分割的精度與穩健性，以應對複雜且多變的真實場景挑戰。例如，在光線條件變化劇烈或使用者快速頭部運動的情況下，系統的穩定性仍有進一步優化的空間。透過改良演算法並採用更精密的模型，我們希望系統能在更極端的環境中維持高效能。

此外，我們將探索降低計算複雜度的方法，以提升系統的運行效率，特別是在硬體資源有限的裝置上實現流暢的實時運行能力。這不僅將使技術在手持裝置和可穿戴設備上更具應用價值，還能擴大其在日常生活中的使用場景。同時，深度學習技術的快速發展為本研究帶來了新的可能性。我們計劃整合更先進的深度學習模型，進一步增強系統在動態背景下或多用戶環境中的表現，從而提高系統的智能程度與應用廣度。

除了技術層面的改進，我們也將積極推動視線追蹤技術在跨領域的應用，包

括虛擬實境（VR）與增強現實（AR）等新興科技領域。在這些領域中，視線追蹤技術能夠顯著提升人機互動的自然性與沉浸感。此外，該技術在醫療輔助診斷、行為分析與教育科技中的應用也具有極高的潛力，例如用於分析癲癇患者的視線行為或提升學生在數位教室中的專注度監測。

為了確保技術的實際可行性，我們計劃在更多真實場景中進行測試，例如駕駛輔助系統、教育訓練平台與娛樂互動應用。這些測試不僅能驗證系統在日常使用中的可靠性，也有助於我們進一步改進系統設計以滿足不同應用場景的需求。透過這些努力，我們期望此技術能夠更廣泛地被接受並應用，為未來的人機互動帶來全新的體驗與價值。

總而言之，本研究為視線追蹤技術的發展奠定了堅實的基礎，未來的研究將在技術優化、應用拓展與實際測試中持續深入，致力於推動該技術在多個領域的創新發展。我們相信，隨著相關研究的進一步推進，此技術將在智慧科技領域中發揮越來越重要的作用，並為社會創造更多價值。

References

- [1] X. Wang, J. Zhang, H. Zhang, S. Zhao, and H. Liu, "Vision-based gaze estimation: A review," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 14, no. 2, pp. 316–332, 2021.
- [2] E. Arcoverde, R. Duarte, R. Barreto, J. Magalhaes, C. Bastos, T. Ing Ren, and G. Cavalcanti, "Enhanced real-time head pose estimation system for mobile device," *Integrated Computer Aided Engineering*, vol. 21, pp. 281–293, 04 2014.
- [3] N. Ruiz, E. Chong, and J. M. Rehg, "Fine-grained head pose estimation without keypoints," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 2074–2083.
- [4] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3d faces," in *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, 2023, pp. 157–164.
- [5] A. K. Chaudhary, R. Kothari, M. Acharya, S. Dangi, N. Nair, R. Bailey, C. Kanan, G. Diaz, and J. B. Pelz, "Ritnet: Real-time semantic segmentation of the eye for gaze tracking," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. IEEE, 2019, pp. 3698–3702.
- [6] Z. Wang, J. Chai, and S. Xia, "Realtime and accurate 3d eye gaze capture with dcnn-based iris and pupil segmentation," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 1, pp. 190–203, 2021.
- [7] S.-J. Baek, K.-A. Choi, C. Ma, Y.-H. Kim, and S.-J. Ko, "Eyeball model-based iris center localization for visible image-based eye-gaze tracking systems," *IEEE Transactions on Consumer Electronics*, vol. 59, no. 2, pp. 415–421, 2013.
- [8] W.-C. Kao, K.-J. Huang, and Y.-C. Chiu, "Eyeball model construction with head

movement compensation for gaze tracking systems,” in *2020 IEEE International Conference on Consumer Electronics (ICCE)*, 2020, pp. 1–2.

- [9] W.-C. Kao, J.-Y. Li, S.-C. Lin, and Y.-C. Chiu, “High precision canthus alignment for visible-spectrum gaze tracking system,” in *2019 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-TW)*, 2019, pp. 1–2.
- [10] R. Valle, J. M. Buenaposada, and L. Baumela, “Multi-task head pose estimation in-the-wild,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 8, pp. 2874–2881, 2020.
- [11] Y. Feng, H. Feng, M. J. Black, and T. Bolkart, “Learning an animatable detailed 3D face model from in-the-wild images,” *ACM Transactions on Graphics, (Proc. SIGGRAPH)*, vol. 40, no. 8, 2021. [Online]. Available: <https://doi.org/10.1145/3450626.3459936>
- [12] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.
- [13] G. Huang, Z. Liu, and K. Q. Weinberger, “Densely connected convolutional networks,” *CoRR*, vol. abs/1608.06993, 2016. [Online]. Available: <http://arxiv.org/abs/1608.06993>
- [14] S. J. Garbin, Y. Shen, I. Schuetz, R. Cavin, G. Hughes, and S. S. Talathi, “Openeds: Open eye dataset,” *arXiv preprint arXiv:1905.03702*, 2019.
- [15] F. N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally, and K. Keutzer, “Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <1mb model size,” *CoRR*, vol. abs/1602.07360, 2016. [Online]. Available: <http://arxiv.org/abs/1602.07360>

- [16] K. Fornalczyk and A. Wojciechowski, “Robust face model based approach to head pose estimation,” in *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)*, 2017, pp. 1291–1295.
- [17] H. Proenca, S. Filipe, R. Santos, J. Oliveira, and L. A. Alexandre, “The ubiris.v2: A database of visible wavelength iris images captured on-the-move and at-a-distance,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pp. 1529–1535, 2010.
- [18] M. Arsalan, R. A. Naqvi, D. S. Kim, P. H. Nguyen, M. Owais, and K. R. Park, “Irisdensenet: Robust iris segmentation using densely connected fully convolutional networks in the images by visible light and near-infrared light camera sensors,” *Sensors*, vol. 18, no. 5, 2018. [Online]. Available: <https://www.mdpi.com/1424-8220/18/5/1501>



自傳

我是陳璽文，來自新竹縣，一個自然而純樸的地方。我的性格內斂穩重，喜歡深思熟慮後行動，這樣的特質也影響了我對學術與人生規劃的態度。從小在父母的教導下，我養成了謹慎而不失開放的心態，並學會持續追求自我成長。他們總是告訴我：「學習不只是為了成績，而是為了開拓更多的可能性。」這句話深深影響了我，成為我面對挑戰的動力來源。

我於高中畢業後，考入國立中央大學數學系。起初，我對純理論的數學懷有極大的興趣，然而隨著課業的深入，發現自己在抽象理論的學習上進展較為緩慢。儘管成績並不拔尖，但我始終抱持著對未知領域探索的熱情，持續尋找數學與實際應用的連結。

大三時，我的人生出現了轉折點。在修讀陳建隆教授開設的「數學影像處理」課程時，我第一次接觸到結合程式設計與數學的應用領域。課程內容深入淺出，讓我明白原來數學不僅僅是紙上的理論，還可以成為解決實際問題的有力工具。這門課啟發了我對程式設計的興趣，也讓我開始探索如何將數學模型應用於影像分析、機器學習等領域。

為了實踐這份興趣，我積極參與跨領域的產學合作計畫。在這些計畫中，我與來自不同背景的同學合作，學習如何將數學技術應用於實際問題，例如影像辨識與數據處理。這些經驗不僅提升了我的技術能力，也讓我更加清楚自己的志向——投入電機資訊領域，發掘數學與科技融合的無限可能。

在確立目標後，我努力準備研究所考試，並成功進入國立台灣師範大學電機研究所進修。在這裡，我希望能繼續深化對人工智慧、影像處理與資料分析的研究，並將數學邏輯與電機技術相結合。我相信，紮實的數學基礎加上程式應用能

力，將使我在電資領域中具備獨特的優勢。

未來，我的目標是成為一名能將理論與實務完美結合的專業工程師，專注於智慧科技的開發與創新。我希望能參與更多產學合作計畫，將所學的知識轉化為對社會有益的技術，甚至進一步貢獻於教育，讓更多人了解數學的魅力與價值。

從數學到程式設計的轉變，讓我明白熱情與實踐的力量。每一段經歷都是一次成長，而每一次成長都讓我離夢想更近一步。我感謝每位在求學路上啟發我的老師與夥伴，也期許自己能在未來不斷突破自我，成為一個對社會有貢獻的人。



學術成就

論文發表

W. -C. Kao, X. -W. Chen and J. -L. Chern, "Gaze Tracking with Head Pose Estimation and Compensation," 2025 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA

