

適用於線上會議之虛擬替身臉部姿態軟體開發

3D Avatar software for video conference

指導老師:黃朝宗

組別: B240

組員:林孟平

Abstract

在疫情時代下，為了順暢地進行居家工作，人們經常使用Google meet、Teams等軟體進行線上視訊會議。然而，由於人們在家裡工作時通常疏於化妝打扮，因此對於開啟視訊鏡頭難免心存顧忌。因此，我們希望能開發一個能夠在Google meet等會議軟體上輸出虛擬替身臉部姿態的軟體，此替身能夠代替人們呈現臉部表情，使得人們在不需在意自身形象的狀況下，放心地開啟鏡頭和其他會議參與人互動。

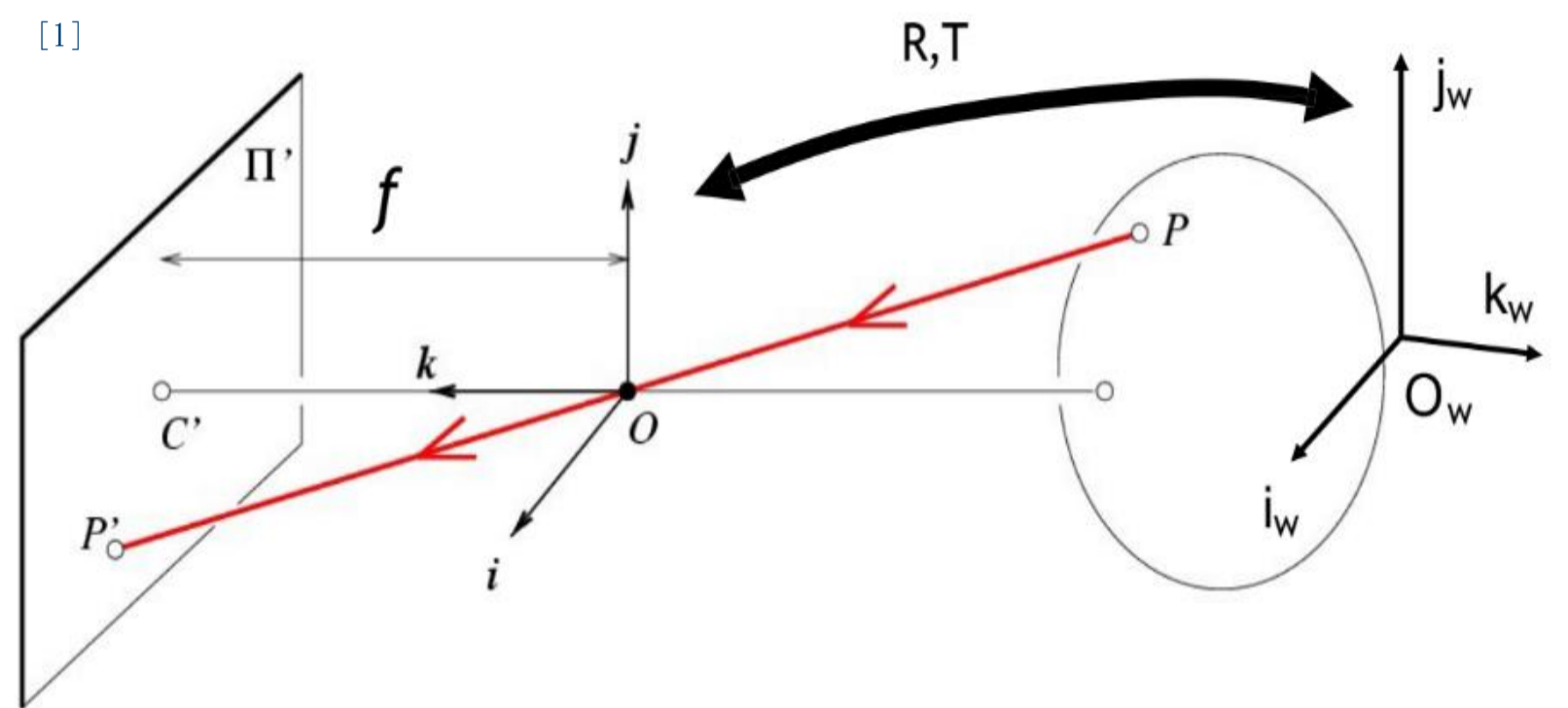
為了讓電腦學習辨識使用者的臉部姿態，我們使用神經網路偵測臉部特徵點，並使用Pinhole camera model之數學模型，除了預測特徵點的3D座標以辨識臉部表情，也計算頭部相對於鏡頭的旋轉角度，以驅動在建模軟體設計之虛擬替身。最後在會議軟體上建立虛擬攝影機來源，搭建出可於會議軟體上運行之系統。

Methodology

- (1) 在會議進行中，假設頭部幾何中心大致呈現不動，可將鼻子定義為世界參考座標系之原點，且知頭部各器官於此座標系之位置：

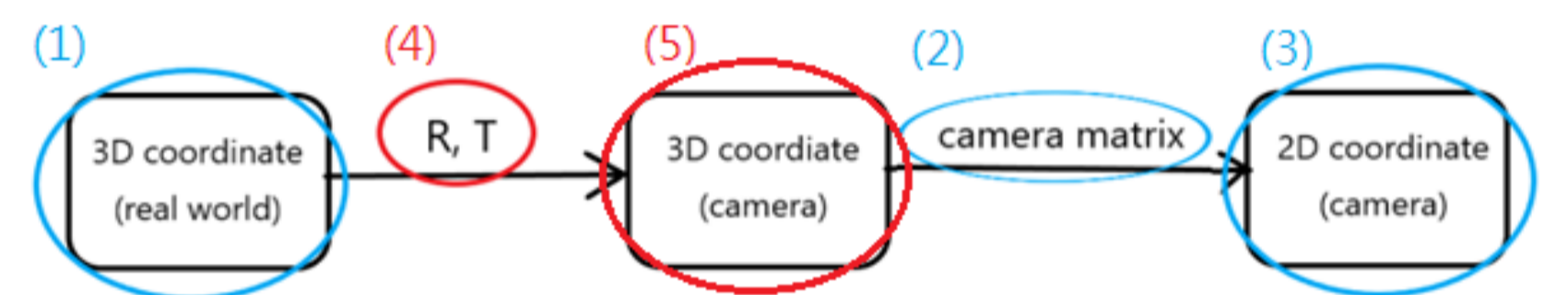
Organ	Px	Py	Pz
Nose	0	0	0
Chin	0	-300	-50
Left Eye	200	150	-130
Right Eye	-200	150	-130
Mouth (Left corner)	150	-150	-130
Mouth (Right corner)	-150	-150	-130

(Unit: Pixel)



- (2) 由於頭部幾何中心不動，則已知鏡頭焦距，得相機矩陣K，此矩陣正是代表相機座標系中2D與3D座標間之轉換關係：

$$K = \begin{bmatrix} \alpha & 0 & 0 & 0 \\ 0 & \alpha & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad P' = KP \quad P' = [x', y', 1]^T \quad P = [x, y, z, 1]^T$$



- (3) 透過神經網路偵測，可得各器官位於相機坐標系之2D座標：

$$P_i = P_i(x, y)$$

已知相機座標系之3D座標與旋轉矩陣後，可知以下控制參數：

Head Rotation Angle

Eye (Mouth) Aspect Ratio

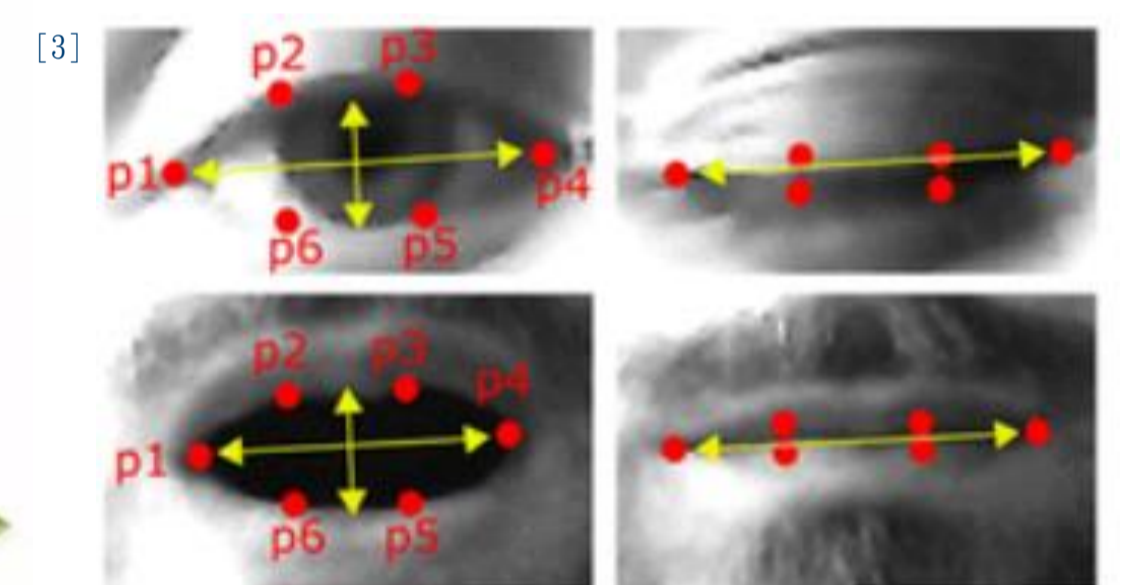
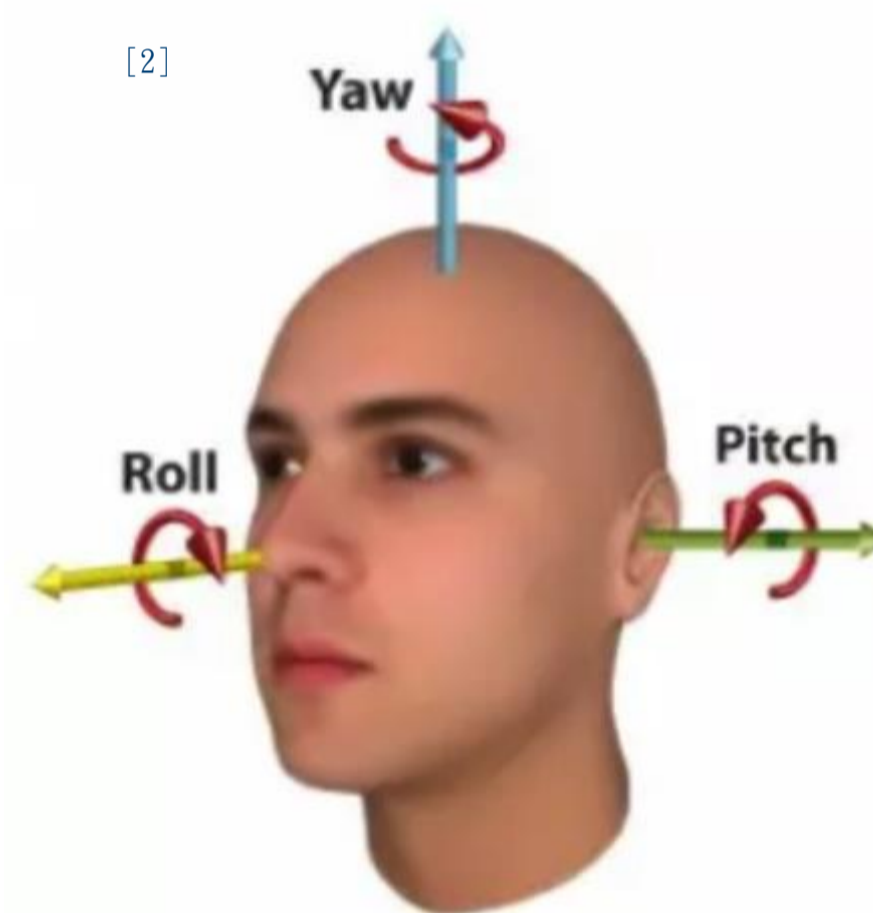
- (4) 使用SolvePnP之方法，可得姿態矩陣M，此姿態矩陣正是代表相機座標系與世界參考座標系間之轉換關係：

$$M = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \quad P^c = MP^w \quad P^c = [x^c, y^c, z^c, 1]^T \quad P^w = [x, y, z, 1]^T$$

- (5) 將M矩陣對(1)設定之座標作用後，可得相機座標系之3D座標，並將其中之R矩陣分解，可分別得到於X, Y, Z方向之旋轉矩陣：

$$R = R(\alpha)R(\beta)R(\gamma)$$

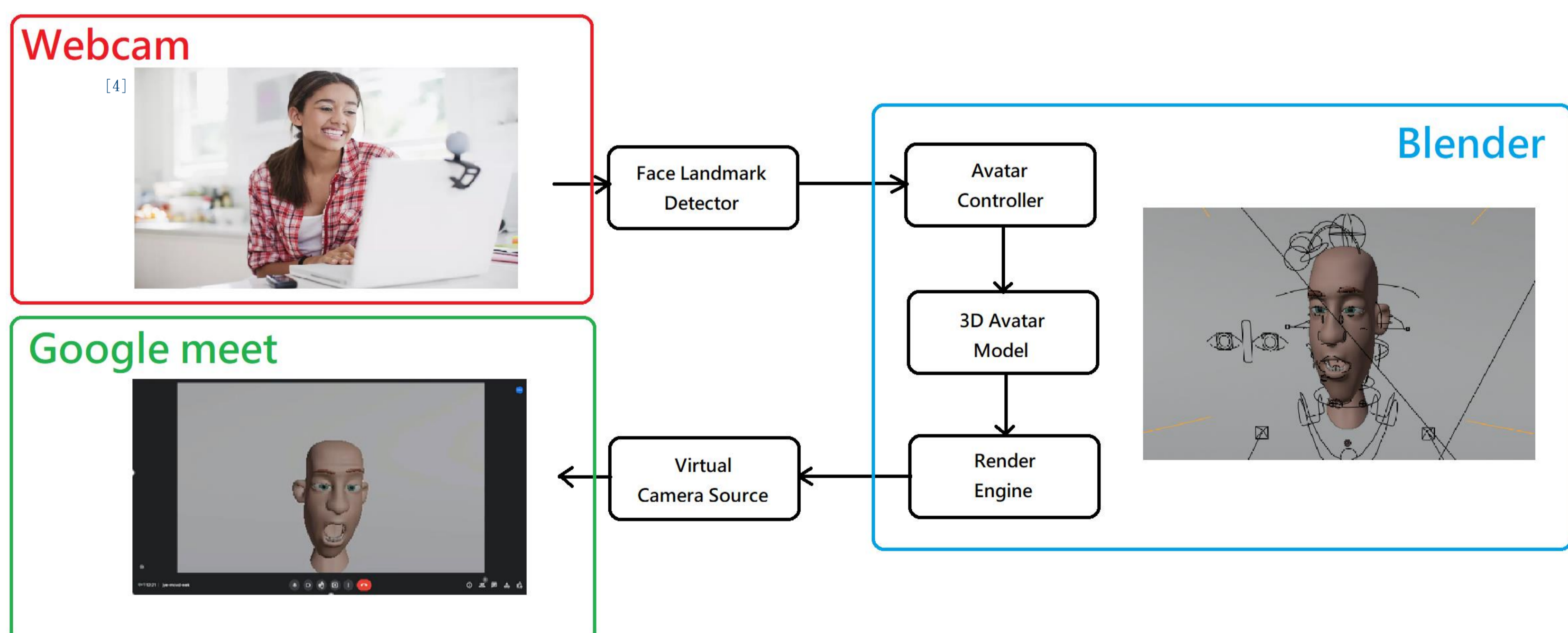
$$R(\alpha) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & -\sin(\alpha) \\ 0 & \sin(\alpha) & \cos(\alpha) \end{bmatrix} \quad R(\beta) = \begin{bmatrix} \cos(\beta) & 0 & \sin(\beta) \\ 0 & 1 & 0 \\ -\sin(\beta) & 0 & \cos(\beta) \end{bmatrix} \quad R(\gamma) = \begin{bmatrix} \cos(\gamma) & -\sin(\gamma) & 0 \\ \sin(\gamma) & \cos(\gamma) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$



$$EAR = MAR = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|}$$

Reference :
Image [1] https://drive.google.com/file/d/1Glp15bubg_tmhameE60sgT0azVBN_ng/view
Image [2] <https://ithelp.ithome.com.tw/articles/10239600>
Image [3] https://www.researchgate.net/figure/The-eye-aspect-ratio-is-practically-identical-to-the-mouth-aspect-ratio_fig22_323953620

System Overview



Reference :
Google Mediapipe Face Mesh Solutions : https://google.github.io/mediapipe/solutions/face_mesh.html
Blender : <https://www.blender.org/>
Open Broadcast Software : <https://obsproject.com/>
Avatar Model : <https://studio.blender.org/characters/5718a967c379cfd04929a4247a1/>
Image [4] : <https://www.digitalcameraworld.com/buying-guides/the-best-webcam-for-home-working>

Conclusion

從結果顯示，使用神經網路進行2D特徵點偵測，並使用電腦視覺之數學方法還原3D特徵點位置及計算頭部旋轉尤拉角，可有效地操縱虛擬替身之臉部姿態，並將此作為會議軟體的影像輸出。尤於現今之神經網路技術已可同時辨識影像中多個人像之特徵點，若是我們能另行開發一個模組先行預測使用者的頭部幾何中心之位置，則不僅可修正人臉不在鏡頭正前方時，座標估計較為不準確的問題，還能一次估計多個使用者的臉部姿態，讓多個虛擬替身在線上會議中顯現。