

從單張單眼人臉影像建立 3D 人臉模型

Reconstruction of a 3D Facial Model from Monocular Facial Images

何信瑩(Shinn-Ying Ho)* 黃慧玲(Hui-Ling Huang)

逢甲大學資訊工程系

Department of Information Engineering, Feng-Chia University, Taichung, Taiwan.

摘要

本文在探討自動建立虛擬實境中人臉的 3D 模型。本文提出一套以人臉為中心的 3D 人臉重建演算法,透過單張或多張 2D 人臉影像來自動修正事先建立好之一般型 3D 人臉模型。本系統可利用一套被動式電腦視覺技術快速計算出影像中人臉的 3D 方位,並依據人臉特性知識為基底,利用影像中各器官特徵點來重建 3D 模型的重要特徵點。配合一般 3D 人臉模型的三角化頂點間的關係,可快速建立特定人的 3D 人臉模型。其特點一在於每張人臉影像的拍攝視角不限定正面且不必事先校正;特點二是若有多張影像則各張影像之間的視角關係可未知,且越多張影像可得到愈精確的人臉模型。配合所提供之二維人臉影像的紋路及貼圖技術,可便利地建立虛擬的 3D 人臉。

關鍵字:電腦視覺、虛擬實境、人臉模型、模型重建

Abstract

In this paper, we investigate the reconstruction of a 3D facial model from monocular facial images in an image-based virtual reality (VR) system. An efficient approach for reconstructing a face-centered 3D facial model is presented. The system creates a generic 3D facial model first and then stores the depth-based knowledge of this facial model. From one or more monocular facial images whose interrelationships are not known, a passive vision-based technique is used to determine the pose of the human face and then the estimated facial features are represented by the transformed face centered coordinates. A specific 3D facial model can be established using the estimated face-centered features to model the generic model based on the depth-based knowledge. The merits of our method are (a) the viewing angles of the facial images are not limited to some specific directions and (b) the more accurate facial model can be obtained if more facial

This work was supported by the National Science Council, Taiwan, ROC. Grant No. NSC 87-2213-E035-014.

*Corresponding author. E-mail: syho@fcu.edu.tw

images are used. The 3D human face in an image-based VR system using the reconstructed facial model with texture mapping.

Keyword: Computer Vision, Virtual Reality, Facial Model, Model Reconstruction

1.前言

重建 3D 表面結構在電腦視覺應用上是一個重要的課題。在稍早的文獻中,大都採用 $2\frac{1}{2}$ D 影像為中心(image-centered)的重建法,如用距離影像(range image)[1]、雙眼影像(stereo)[2],陰影影像[3],另外尚有採用混合二種資源以上的方法[4-5]。

近來,採用 3D 表示法開始被重視,例如採用 3D 表面綴面法[6],參數表面法[7-8]。這些以 3D 表示的做法大都採用單張影像的資訊,如邊緣(silhouettes)和影像特徵[9],採用雙眼影像(stereo)[10]或距離影像[11]來修正事先建好的估計模型。

本文所要探討的主題是建立虛擬實境(VR)系統所需要的虛擬實物,其特點有

- (1)必須能以輸入影像以外的視角呈現,因此要能以 3D 表示法來建立模型。
- (2)模型呈現時,將以紋路對映(texture mapping)技術對模型做貼圖,以表現真實感。紋路來源為建立模型所取用影像的紋路。
- (3)虛擬實物的構建尤以 3D 人臉模型重建為重,特別是建立特定人的 3D 模型。現有技術,如採用雷射光掃描儀,3D 數位量測儀都需要真實 3D 人臉,做法較不方便。

Wallace[12]提出以一張正面和一張側面且已校正好方位的二張影像來建立 3D 人臉模型。配合上述虛擬實物特點,我們需要能簡易快速建立 3D 人臉。本論文提出一套建立 3D 人臉模型的演算法,其特色如下:

- (1)採用以物體為中心(object-centered)的 3D 表示法,是以包含特徵點為頂點(vertex)的三角化 3D 綴面所組成的一般型(generic)人臉模型。
- (2)輸入影像為 2D 單張單眼的黑白或彩色影像,而比較複雜計算的立體像或距離影像。本文提出以深

度為基礎(depth-based)方法來估算 2D 影像失去的深度資料。

- (3) 重建 3D 模型是以修正粗略模型,修正方法是最小化目標函數。若有多張影像,其彼此視角關係未知的影像,則可逐步重覆調整(incremental refinement)3D 模型。
- (4) 採用被動式電腦視覺(passive vision-based)技術,取用單張影像的內部紋路及特徵點資訊來重建模型,人臉不必貼任何標記,拍攝影像不必事先校正,也不必額外的儀器輔助,可使本方法更具實用性。3D 人臉模型重建系統流程圖請參照圖一。

2. 構建一般型人臉模型與特徵點擷取

2.1 人臉模型構建

為配合貼圖及易於調整 3D 人臉模型,一般型人臉模型構建要點如下:

- (1) 以人臉為中心(face-centered)的三角化綴面表示法。如圖二所示,令人臉模型為左右對稱,眼外角 P_1 及 P_2 , 和嘴角 P_3 及 P_4 四點所成平面為 XY 平面, 且 $\overline{P_1P_2} \parallel \overline{P_3P_4}$ 。令 P_5 和 P_6 為 $\overline{P_1P_2}$ 和 $\overline{P_3P_4}$ 的中點, P_7 為 $\overline{P_5P_6}$ 的中點。令 P_7 為座標原點, $\overline{P_7P_5}$ 為正 Y 軸, $\overline{P_1P_2}$ 為 X 軸, 平面法向量向外為正 Z 軸。人臉模型的透視投影幾何關係圖可參考圖三。
- (2) 三角化綴面的頂點(vertex)選取原則是以人臉的眼、眉、唇和鼻等在影像平面上明顯特徵點優先選取,另外以正面人臉邊緣(silhouette)與人臉對稱軸、眼水平線、口角水平線的交點等特徵點選取為頂點。上述頂點稱為特徵點(feature vertices)。這些特徵點的主要功能為人臉內部資訊點的定位和人臉外形的修正。然後再以 Hermite 參數曲面對一般型 3D 人臉模型建立一些非特徵點的頂點。
- (3) 記錄每一個非特徵點和鄰近特徵點的相對方位關係。考慮每一頂點 (x, y, z) 在 XY 平面的平行投影,即只考慮 (x, y) 值,進行 XY 平面的格子編碼。格子編碼的密度以每格最多有一頂點為原則。此目的有二個功能:一是格子編碼可達正規化的目的,二是易於記錄各頂點之間的方位關係。此外,記錄個各相鄰頂點之 Z 值的比例關係,這資訊將提供以深度為基礎(depth-based)的重建技術估算 2D 影像所失去的深度資料。
- (4) 一般建立人臉模型約需數千個頂點,然而我們的人臉模型會配合貼上人臉紋路呈現,因此可大大降低頂點數目圖四所示之一般型人臉模型共用 527 個頂點。貼圖技術可參考 [13,14]。

2.2 人臉特徵點擷取

自動從單張影像擷取眼、眉、鼻、唇...等特徵是一個值得研究的課題。本系統將採用人工輔助方

式選取所需特徵點。若欲自動選取特徵點,其技術可參考[15]

3. 計算影像人臉 3D 方位

欲從單張單眼影像重建影像中人臉的 3D 模型並不容易,除了缺乏深度資料外,影像中人臉的方位並不受固定方向約束是困難所在。本節提出一套演算法可快速測定人臉方位,再利用影像特徵去重建粗略 3D 模型,在下一節說明如何利用已建好之一般型模型來修正。

3.1 人臉 3D 方位測定演算法

從單張單眼影像去做 3D 人臉方位的測定技術可參考我們最近發表的文章 [16]。今摘要說明如下。

給定人臉為中心之臉部模型(其定義與本文相同)及已知兩眼距離。可取得二眼角和二嘴角共四個特徵點的情況下,以下演算法可快速唯一決定 3D 模型的方位。

步驟一:由投影中心 C 點和影像上線段 $\overline{P_1P_2}$ 和 $\overline{P_3P_4}$,

利用 $\overline{P_1P_2} \parallel \overline{P_3P_4}$ 特性可決定眼線向量 N_e 。

$$n_p = (x_1, y_1, f) \times (x_2, y_2, f)$$

$$n_q = (x_3, y_3, f) \times (x_4, y_4, f)$$

$$N_e = \frac{(n_p \times n_q)}{\|n_p \times n_q\|}$$

步驟二:由投影中心,二眼角連線向量(等於 X 軸向量)及二眼距離 L_e 可得 3D 點 P_1, P_2 及對稱軸點 P_5 。

$$(U, V, W) = L_e \cdot (u, v, w)$$

$$X_i = \frac{x_i}{f} Z_i, \quad Y_i = \frac{y_i}{f} Z_i,$$

$$Z_i = \begin{cases} \left(\frac{(fU - Wx_{3-i})}{(x_{3-i} - x_i)} \right), & \text{if } x_{3-i} \neq x_i \\ \left(\frac{(fV - Wy_{3-i})}{(y_{3-i} - y_i)} \right), & \text{if } y_{3-i} \neq y_i, \end{cases}$$

$$i \in \{1, 2\}.$$

$$(X_s, Y_s, Z_s) = \frac{1}{2} ((X_1+X_2), (Y_1+Y_2), (Z_1+Z_2)).$$

步驟三:利用 $\overline{P_5P_6}$ 為對稱軸且與 $\overline{P_3P_4}$ 垂直,可得 3D 點 P_6, P_3 和 P_4 。令

$$Z_{abs}(i) = \begin{cases} \left(\frac{(fu - wx_{7-i})}{(x_{7-i} - x_i)} \right), & \text{if } x_{7-i} \neq x_i \\ \left(\frac{(fv - wy_{7-i})}{(y_{7-i} - y_i)} \right), & \text{if } y_{7-i} \neq y_i. \end{cases}$$

$$X_6 = \frac{L_m}{2f} [x_3 Z_{abs}(3) + x_4 Z_{abs}(4)] = L_m C_x,$$

$$Y_6 = \frac{L_m}{2f} [y_3 Z_{abs}(3) + y_4 Z_{abs}(4)] = L_m C_y,$$

$$Z_6 = \frac{L_m}{2} [Z_{abs}(3) + Z_{abs}(4)] = L_m C_z,$$

$$\text{因 } \overline{P_5 P_6} \cdot N_c = 0 \text{ 則 } L_m = \frac{(X_5, Y_5, Z_5) \cdot N_c}{(C_x, C_y, C_z) \cdot N_c}$$

$$X_i = \frac{x_i}{f} Z_i, \quad Y_i = \frac{y_i}{f} Z_i,$$

$$Z_i = \begin{cases} L_m \cdot \left| \frac{(fu - wx_{7-i})}{(x_{7-i} - x_i)} \right| & \text{if } x_{7-i} \neq x_i \\ L_m \cdot \left| \frac{(fv - wy_{7-i})}{(y_{7-i} - y_i)} \right| & \text{if } y_{7-i} \neq y_i, \end{cases}$$

$$i \in \{3, 4\}$$

步驟四: $N_s = \frac{\overline{P_5 P_6}}{\| \overline{P_5 P_6} \|}$ 爲人臉對稱軸向量而 $N_f = N_c \times N_s$ 爲人臉法向量。

測試一標準圖(Lena) 因相機的參數本質不知道, 用一般常識評估焦距 f 。所得人臉單位向量軸如下:

$$\text{眼線}(N_{ex}, N_{ey}, N_{ez}) = (-0.8375, -0.0050, -0.5491)$$

$$\text{對稱軸}(N_{sx}, N_{sy}, N_{sz}) = (-0.0657, -0.9919, 0.109)$$

$$\text{人臉平面法向量}(N_{fx}, N_{fy}, N_{fz}) = (0.5452, -0.1272, -0.8286)$$

可得兩個模型比率 $L_m/L_c = 0.5207$ 反投影的結果建立 3D 模型方位顯示於圖二。

3.2 人臉方位測定演算法之誤差分析

本節將考慮特徵點因影像雜訊及擷取誤差對演算法所造成之誤差的敏感度分析。由於透視投影函數是非線性, 較難以解析式分析, 我們將以實驗式的模擬分析各個人臉方位下, 其人臉法向量 N_f 和觀測距離 ($\| \overline{CP_i} \|$) 的誤差。這些資訊將提供修正 3D 模型頂點時, 做爲誤差容忍度參考。

令從 2D 人臉影像所測得法向量 $N_2 = (\sin \sigma \cos \tau, \sin \sigma \sin \tau, -\cos \sigma)$ 及觀測距離爲 D , 其中 σ 爲 N_2 和光軸的夾角, τ 爲 N_2 在 xy 投影平面之投影向量和 x 軸之夾角。由於人臉對稱, 故 τ 只是分析 -90° 至 90° 即可, 而在取得四眼嘴角特徵點下, σ 約在 0 至 30°

範圍。以經驗值判定, 當近距離拍攝人臉影像時扭曲效果較大。以下實驗令觀測距離爲人眼寬 5 倍, 而 p_1 因雜訊取得 p_1' , 令 $p_1' p_2$ 和 $p_1 p_2$ 之夾角爲 0.5° , 而 p_2, p_3 和 p_4 不變情況下, 測得觀察距離爲 D' , 而人臉法向量 N_f' 。圖五和圖六分別顯示在不同的 σ 和 τ 方位下, 距離誤差 $E_D = \frac{|D - D'|}{D} \times 100$ 和法向量誤差

$E_N = \cos^{-1}(N_f \cdot N_f')$ 。由實驗分析可知 τ 在靠近 $\pm 90^\circ$ 時誤差較大, 而 σ 愈大 (即愈偏離正面) 誤差愈大, 若提供額外事先學習的資訊, 如眼線與嘴巴距離, 對雙眼距離之比例, 可大爲降低誤差。

4 修定 3D 人臉模型

欲從單張單眼 2D 影像去評估特徵點的深度資訊並不是一件直接而簡單的事。本文將以深度爲基礎的技術, 從一般型模型所記錄特徵點離人臉平面的距離和特徵點在人臉平面上投影位置資訊, 來推算每一特徵點的 3D 位置。

4.1 估計特徵點的初始位置

令特徵點 P_i 在人臉爲中心座標系統中座標爲 $P_{fi} = (x_{fi}, y_{fi}, z_{fi})$ 。如圖七所示, 假設由上一節所敘述方法測得人臉平面方程式爲 $ax + by + cz = D$, 其中 $N_f = (a, b, c)$ 。在滿足 P_i 的深度資訊爲 z_{fi} 下, 由 2D 影像特徵點 $p_i = (x_i, y_i)$ 可得估算的特徵點 p_i' , 其相對觀察座標系統的座標爲 (x_{ti}, y_{ti}, f_{ti}) , $t_i = (D - z_{fi}) / (ax + by + cz)$ 。

將 P_i' 經由平移及旋轉矩陣轉換爲相對人臉爲中心之座標。

$$\begin{bmatrix} x_{fi} \cdot w \\ y_{fi} \cdot w \\ z_{fi} \cdot w \\ w \end{bmatrix} = \begin{bmatrix} N_{ex} & N_{ey} & N_{ez} & 0 \\ N_{sx} & N_{sy} & N_{sz} & 0 \\ N_{fx} & N_{fy} & N_{fz} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & -x_7 \\ 0 & 1 & 0 & -y_7 \\ 0 & 0 & 1 & -z_7 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_{ti} \\ y_{ti} \\ f_{ti} \\ 1 \end{bmatrix}$$

4.2 調整 3D 特徵點位置

依據估算的 P_i' 值, 我們設定目標函數值 (objective function) $f(P)$ 來調整所有特徵點 P_{fi} 爲 P 點。注意 P_i' 值爲參數 t_i 的函數值, 而 t_i 爲 P_{fi} 深度的函數值。令 $f(P)$ 函數定義如下:

$f(P) = w_1 f_1(P) + w_2 f_2(P)$ 其中 w_1 和 w_2 爲加權系數。本系統採用動態規劃 (dynamic programming) 方法計算 P 點位置, 使得 $f(P)$ 爲最小值。定義 $f_1(P)$ 是調整特徵點 P_{fi} 爲 P 後, P 點在觀察座標的投影點和 2D 影像特徵點的差距函數, P 的投影點愈靠近 p_i 即 P_{fi} 調靠近 $\overline{CP_i}$, 如此可最小化 $f_1(P)$ 。 $f_2(P)$ 是各 P 點的彼此關係

與事先學習的人臉特徵點彼此關係的差距函數。彼此關係包含深度比例,在人臉平面投影點位置關係及符合人臉特色,如對稱等關係。

4.3 修正 3D 模型的三角化頂點

在 2D 影像中只能擷取特徵點,而無法找到非特徵點之頂點投影,故無法用 2D 影像調整 3D 的非特徵點。本系統做法是利用事先學習所得非特徵點與鄰近特徵點之相對方位比例關係。令 u 為非特徵點, f_i ($i=1,2,\dots,n$) 為 u 的 n 個鄰近特徵點, r_i 為學習所得 u 和 f_i 的比例距離, d_i 為調整後的比例距離,調整 u 點使得最小化 $g(u)$ 目標函數。 $g(u) = \sum_{i=1}^n c_i (d_i - r_i)^2$

只有一張影像時,本系統方法可依事先學習的人臉規則估計特定人臉的 3D 模型。但是影像若有遮避人臉特徵點或特徵點選取誤差,或是人臉朝向使得特徵點過少時,都容易造成不夠精確的 3D 模型。本系統希望藉由第二張,甚至多張影像來精確調整 3D 模型。做法可分為二步驟。

步驟一:以人臉為中心(face-centered)座標系統為主,將影像特徵點與 COP 的射線(ray)通過調整後 3D 特徵點記錄,可在下一張影像最小化 $f(P)$ 函數時提供資訊減少 P 的搜尋範圍,因為這些次級特徵點較有可能最小化 $f_i(P)$ 函數。

$$F(P) = \sum_{i=1}^n f_i(P), f_i(P) \text{ 為第 } i \text{ 張影像的目標函數。}$$

步驟二:將先前估算修正過的 3D 人臉模型取代原先一般型模型,使得搜尋範圍變小。在修正特徵點時,要最小化 $F(P)$ 函數。

5. 貼圖產生人臉影像

由於模型修正是以特徵點的修正,使模型投影與 2D 影像特徵點重合,因此將非特徵點也投影至 2D 影像便可輕易將 2D 影像三角化,每一個 2D 三角形的紋路都有一個 3D 三角形對應。當某一個三角形有來自兩張以上影像的多張紋路時,可以投影角度來制定有加權係數的紋路對映函數。角度越大,紋路資訊壓縮愈多,加權係數愈小。圖八顯示輸入影像及其建立之 3D 人臉模型及貼圖後結果。

6. 結論

本論文提出一套配合虛擬實物構建的方便演算法,而不採用傳統建立精確 3D 模型的演算法,因為本方法是搭配紋路貼圖達到另一視角成像時仍有高度真實性。其特色如下:

- (1) 採用以人臉為中心(face-centered)的 3D 表示法,是以特徵點為頂點(vertex)的三角化 3D 綴面所組成的一般型(generic)人臉模型。
- (2) 輸入影像為 2D 單張影像,而非較複雜計算的立體像或距離影像。本文提出以深度為基礎(depth-based)方法來估算 2D 影像失去的深度資

料。

- (3) 若有多張彼此視角關係未知的影像,則可逐步重覆調整(incremental refinement)3D 模型。
- (4) 重建 3D 模型是以修正粗略模型,修正方法是最小化目標函數。
- (5) 不必使用真實 3D 模型,可使用本系統更具實用性。

文獻參考

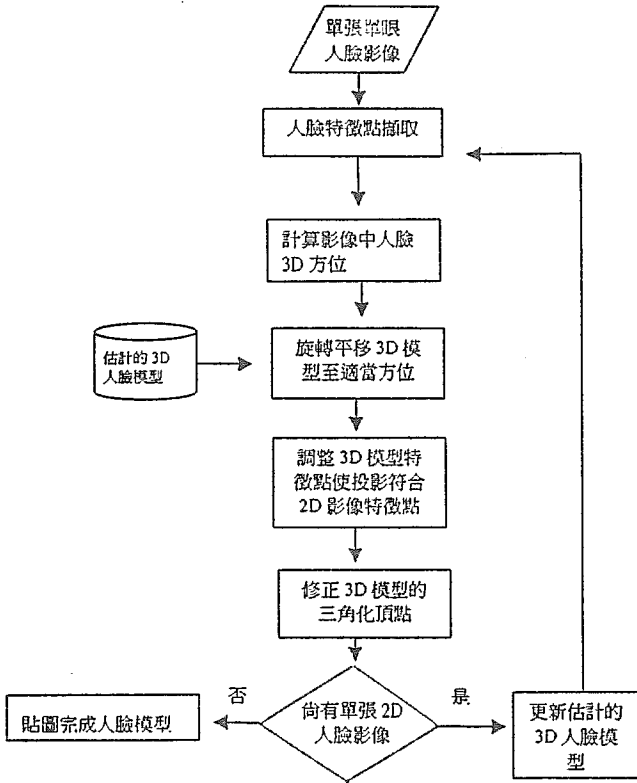
- [1] Asada, M., Kimura, M., Taniguchi, and Shirai, Y. Dynamic integration of height maps into 3D world representation from range image sequences. *Int. Journal of Computer Vision*, 9(1), pp. 31-54, 1994.
- [2] Witkin, A.W., Terzopoulos, D., and Kass, M., Signal matching through scale space. *Int. Journal of Computer Vision*, 1, pp. 133-144, 1987.
- [3] Horn, B.K.P. Height and gradient from shading. *Int. Journal of Computer Vision*, 6(2), pp. 105-132, 1990.
- [4] Choe, Y. and Kashyap, R.L. 3-D shape from a shaded and texture surface image. *IEEE Trans. on PAMI*, 13, pp. 907-919, 1991.
- [5] Abbot, A.L. and Ahuja, N., Active surface reconstruction by integrating focus, vergence, stereo, and camera calibration. In *Int. Conference on Computer Vision*, pp.489-492, 1990.
- [6] Terzopoulos, D. and Metaxas, D., Dynamic 3D models with local and global deformations: Deformable superquadrics. *IEEE Trans. on PAMI*, 13, pp. 703-714, 1991.
- [7] Stokley, E.M. and Wu, S.Y., Surface parameterization and curvature measurement of arbitrary 3-d objects: five practical methods. *IEEE Trans. on PAMI*, 14(8), pp. 833-839, 1992.
- [8] Lowe, D.G., Fitting parameterized 3-D models to images. *IEEE Trans. on PAMI*, 13, pp. 441-450, 1991.
- [9] Cohen, i., Cohen, L.D. and Ayche, N., Introducing new deformable surface to segment 3D images. *Confrence on Computer Vision and Pattern Recognition*, pp..738-739, 1991.
- [10] Fua, P. and Sander, P., Segmenting unstructured 3D points into surfaces. In *European Conference on Computer Vision*, Genoa, Italy, 1992.
- [11] Whaite, P. and Ferrie, F.P., From uncertainty to visual exploration. *IEEE Trans. on PAMI*, 13, pp.10133-1049, 1991.
- [12] Wallace, R.S., Automatic Creation of 3D Facial Models. *IEEE CG&A* pp. 16-22, 1993.
- [13] Shinn-Ying Ho, Li-Sun Shu and Hui-Ling Huang., An efficient texture mapping algorithm based on Newton's divided-difference interpolation polynomials. *Proceedings of the 1997 Int. Conf. CSTIA*, pp. 254-260, 1997.
- [14] Paul S. Heckbert, Pixar, Survey o texture mapping. *IEEE CG&A*, pp.56-67, 1986.
- [15] Chellappa et al.. *Human and Machine Recognition*

of faces: A survey. in Proc. Of th IEEE, Vol.83 No. 5, pp.704-740, 1995.

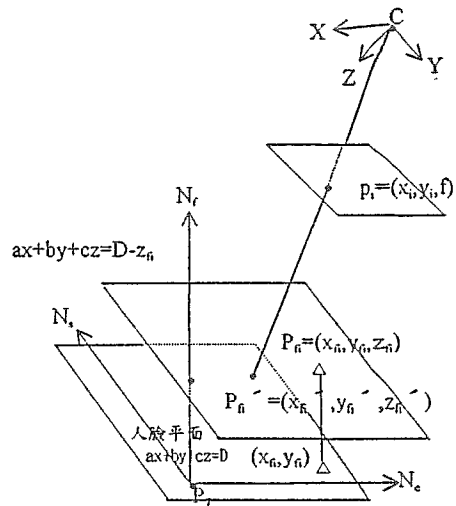
- [16] Shinn-Ying Ho and Hui-Ling Huang, An analytic solution for the pose determination of human faces from a monocular image. Proceedings of 1997 IPPR Conf. on CVGI, pp. 499-506, Taichung Taiwan, Augst 1997.



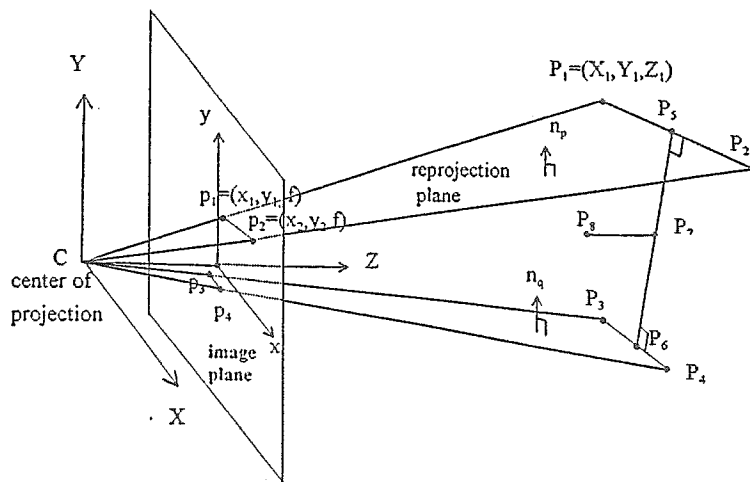
圖二. 3D 人臉幾何模型與人臉方位測定結果



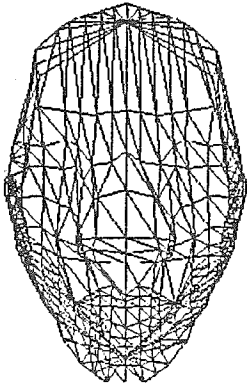
圖一 3D 人臉模型重建流程



圖七. 人臉特徵點估算之幾何參考圖



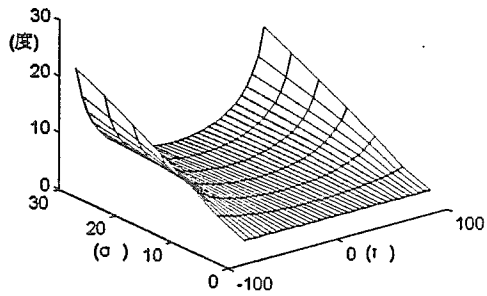
圖三. 反透視投影人臉模型之幾何參考圖。



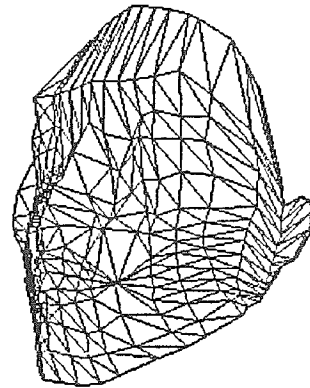
圖四.一般型 3D 人臉模型



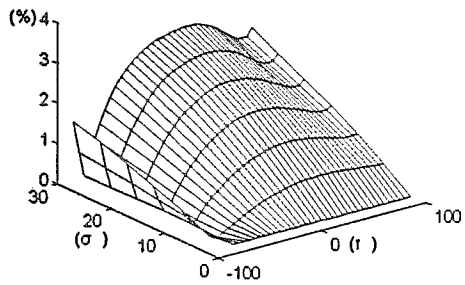
(a)



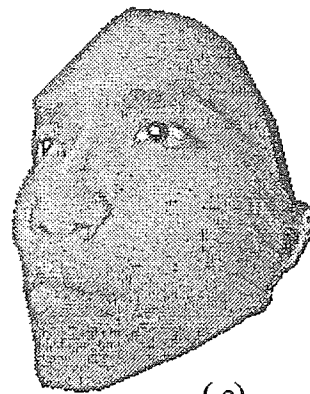
圖五. 不同人臉方位下的距離誤差誤差



(b)



圖六. 不同人臉方位下的法向量誤差



(c)

圖八. 建立虛擬 3D 人臉模型 (a) 待建立模型入影像 (b) 建立好之 3D 人臉模型 (c) 貼圖人臉影像