

魔鏡

■黃建霖 丁俊仁 林基焜 楊家輝 吳宗憲

新一波的數位化浪潮
會帶給消費者新的生活樂趣與娛樂體驗，
結合影像、語音、通訊、網路多媒體等
科技的數位生活，
提供處處為您設想的服務。

在優質數位生活體驗屋中，成功大學電機系及資工系共同研發展示一套多媒體人機互動系統，名為「魔鏡」，它的創意來自於童話故事書《白雪公主》。童話故事中萬能的魔鏡會答覆問題，並提供使用者想知道的資訊。因此，基於便利現代人生活的需求，把日常生活中常用的鏡子創意地設計為也可以提供數位資訊服務。

在科技的運作與連結下，整合網路、通訊、影視娛樂的服務，會讓數位生活就如同身在五星級飯店一樣，無論身在何處，都有處處為你設想的專人生活服務。智慧家電一直是賣場中的熱門商品，這類家用產品提供消費者在影音娛樂上更美好的享受與便利。在世界趨勢及政府的努力推動下，數位家庭概念逐漸在消費者心目中形成，數位生活的市場占有率將逐年提高，而台灣家庭對數位娛樂的需求也會持續增加。

鏡子是生活中常見的用品，也常是童話故事的主角之一。如何結合現代科技與童話故事的想像，是創意的重點。以魔鏡為架構的智慧生活多媒體平台，可實現大家耳熟能詳的「白雪公主」童話故事情節。

神奇的魔鏡

在「優質數位生活體驗屋」中，成功大學電機系及資工系共同研發展示一套多媒體人機互動系統，名為「魔鏡」，它的創意來自於童話故事書《白雪公主》。童話故事中萬能的魔鏡會答覆問題，並提供使用者想知道的資訊。因此，基於便利現代人生活的需





求，把日常生活中常用的鏡子創意地設計為也可以提供數位資訊服務。

魔鏡系統可提供時間查詢、多媒體影音播放、行事曆查詢、網路氣象、新聞檢索等服務。它結合了自動語音辨識技術，以聲控的方式操作魔境，且配合語音合成及立體精靈影像輸出，達到生動便利的人機互動。利用影像追蹤技術，讓魔鏡精靈可以追蹤使用者臉部的位置，擬人化地與使用者交談。

在其他服務方面，也提供像是下載即時的網路新

聞和網路氣象資訊的服務，告訴您最新的生活須知，提供的影音服務可以選播影片和音樂做為休閒用。綜合而言，魔鏡結合了語音訊號處理、影像處理、網路資料檢索等技術，對於未來的數位生活提供了人性化的服務。

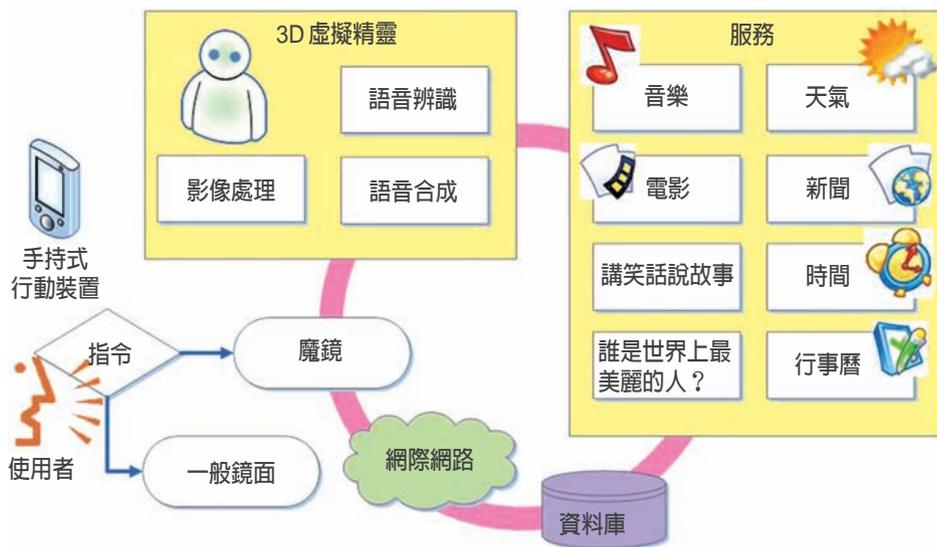
這個創意作品提供的數位生活有無限的想像空間，增加了鏡子及液晶顯示的附加價值，能夠凸顯數位科技對於日常生活的影響。作品如同童話故事「白雪公主」般的情節，擁有魔鏡的所有情境與功能，使

鏡子是生活中常見的用品，也常是童話故事的主角之一。

如何結合現代科技與童話故事的想像，是創意的重點。

以魔鏡為架構的智慧生活多媒體平台，可實現大家耳熟能詳的「白雪公主」童話故事情節。

歷年來，有許多研究把鏡面與顯示器結合，並應用在數位生活上，使得結合鏡面與顯示器所發展的產品擁有無限的想像空間。



魔鏡的功能包括提供一般鏡子和魔鏡精靈的選擇；提供虛擬精靈對話功能；提供語音聲控操作及回應；查詢時間及排定的行程；下載新聞和氣象資訊，並唸給你聽；播放影片及音樂；會追蹤使用者的方向，生動地與使用者交談等。

用的情境舉例如下。

早上起床，小美在浴室梳洗……

小美：魔鏡！魔鏡！

魔鏡：您好！請問需要什麼服務？

小美：來點音樂吧！

魔鏡：為您播放〈……〉

小美：今天天氣如何？

魔鏡：今天天氣多雲時陰短暫陣雨（下載分析網路天氣資料）。

魔鏡：氣溫攝氏 27 度，降雨機率 50%。

小美：有什麼新聞呢？

魔鏡：社會新聞有……（下載分析網路新聞資料）

小美：今天有什麼行程？

魔鏡：您今天上午 10 點，開會。下午 3 點，牙醫。（讀取資料庫內行事曆）

小美：魔鏡！魔鏡！誰是世界上最美麗的人？

魔鏡：當然是您了！親愛的主人。

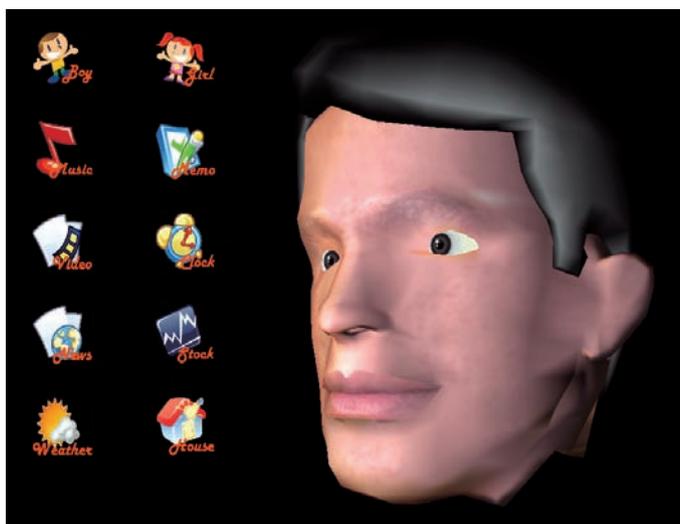
（系統展示可參考 <http://www.youtube.com/watch?V=4CgYQKUoHX4>）

在國際間，數位生活已逐漸成爲一個新的商業市場，很多專家學者完成了許多周延完整的計畫。歷年來，也有許多研究把鏡面與顯示器結合，並應用在數位生活上，使得結合鏡面與顯示器所發展的產品擁有無限的想像空間。

半反射鏡面與液晶顯示器

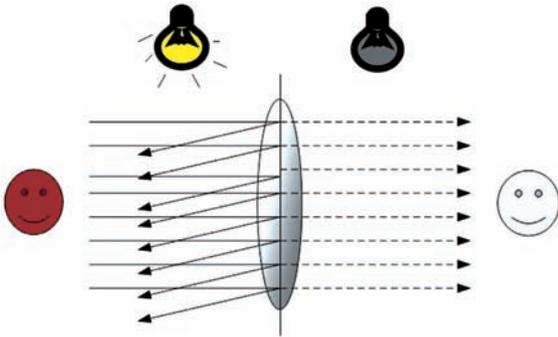
在電視影集和電影裡，常常有證人辨認嫌犯的劇情：證人和犯人會分開在兩個房間，證人可以透過單面反光鏡辨識嫌犯，但嫌犯甚麼都看不見。事實上，單面反光鏡正反兩面的光學性質根本沒有分別，它之所以能產生單面反光的效果，完全是因爲鏡的兩側處於不同光度的環境所造成的。

在實際使用時，嫌犯面向鏡面，而且身處有強光的房間，因爲光線充足，反射的光較多，嫌犯在鏡中便只會看見自己的影像。而證人是站在鏡子的另一邊一個光線很微弱的房間裡，雖然證人房間裡部分的光



魔鏡系統可提供時間查詢、多媒體影音播放、行事曆查詢、網路氣象、新聞檢索等服務。

魔鏡可以透過內建攝影機抓取使用者的臉部影像，在對話中進行影像追蹤，使得 3D 虛擬精靈在對話中能注視著使用者，讓人機互動更為真實。



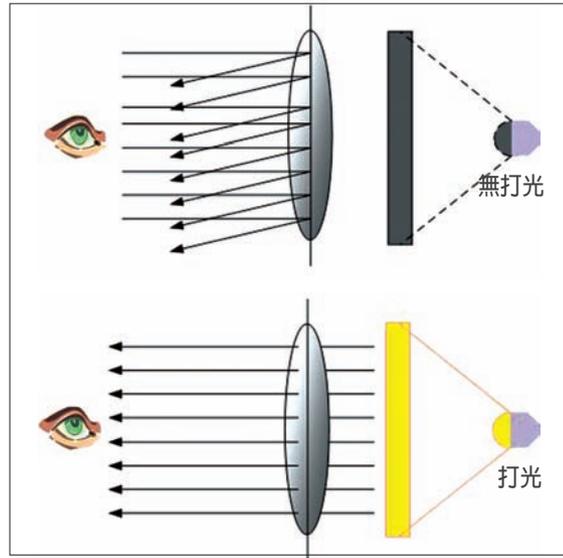
在電視影集和電影裡，常常有證人辨認嫌犯的劇情：證人和犯人會分開在兩個房間，證人可以透過單面反光鏡辨識嫌犯，但嫌犯甚麼都看不見。嫌犯身處有強光的房間（左邊），因為光線充足，反射的光較多，在鏡中便只會看見自己的影像。證人在鏡子的另一邊一個光線很微弱的房間裡（右邊），雖然房間裡部分的光線也可以穿過單面反光鏡，但由於光度很低，因此嫌犯看不到證人。

線也可以穿過單面反光鏡，但由於光度很低，因此嫌犯看不到證人。

運用相同的原理，當液晶螢幕打開時，鏡子前面的人就可以看到液晶螢幕的所有資料。半反射鏡想法類似於電影裡審問嫌犯房間內的鏡子，利用特殊材質，光線明亮的一方可以看見鏡中的反射影像，卻無法看到鏡後方的景物。利用半反射鏡的特性，在鏡後安裝一個影像顯示裝置。當鏡子後方的影像顯示裝置關閉時，整體裝置看起來如同一般鏡子，當鏡子後方的影像顯示裝置投影成像時，使用者便可立即看見成像的資訊內容，以及出現的魔鏡精靈。

自動語音辨識

語音人機介面就是利用語音做為對機器下達命令的輸入媒介，因此首要之務是讓電腦理解人類的語言。語音技術牽涉的範圍極為廣大，包括語言學、音韻學、統計學、聽覺效能、訊號處理等，都是影響語音技術的因素。如何訓練出一組良好的辨識模型，讓它可以模擬各種發音的特性，包含不同人的聲音特質、不同地方的腔調和口音等，都是語音辨識中最重



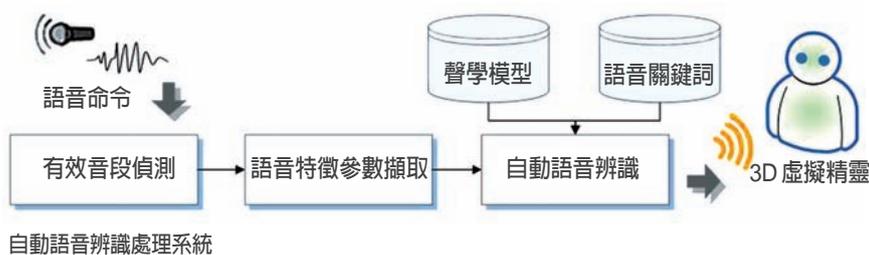
利用半反射鏡的特性，在鏡後安裝一個影像顯示裝置。當鏡子後方的影像顯示裝置關閉時，整體裝置看起來如同一般鏡子。當鏡子後方的影像顯示裝置投影成像時，使用者便可立即看見成像的資訊內容，以及出現的魔鏡精靈。

大的關鍵。語音辨識系統發展已有數十年歷史，但仍有一些問題需要克服。

在早期，研究語音辨識有專家系統、類神經網路等方法，目前全球有關語音研究最常使用的模型，是隱藏式馬可夫模型（hidden Markov model, HMM）。這個模型最早是由馬可夫在 1913 年提出的，發展到今日已成為語音辨識系統的主流，是把語音訊號處理技術應用在人機互動上，最方便自然的方法。

藉由有效音段偵測可以辨別語音輸入的訊號，隨時偵測從音效卡收錄到的聲音。只有正確的交談或命令會被輸入，再估算輸入訊號的能量。由於做為語音輸入的訊號通常會有相對較高的能量，所以需要設定邊界值做為判斷的依據。

在不同的使用場所中，環境聲音對於語音命令訊號的能量會有疊加作用。在較安靜的環境下，如室內，能量通常相對較低。相反地，像是戶外及開車途中較吵雜的環境，語音的能量相對較高。因此在實際應用上，邊界值設定需要隨著環境自動調變，例如



章中的語句做詞的切分，並把字碼轉換成對應的音碼，再透過語音韻律規則資料庫演算出準確的音高、音長、音

利用每次啟動語音命令的前 0.25 秒，偵測環境噪音動態來調整適合的邊界值。

在處理語音訊號時，面對語音這種具有高度變化和差異性的訊號，需要找到穩定且具有鑑別性的特徵參數做為辨識與分類用。魔鏡利用梅爾倒頻譜特徵參數做為語音訊號特徵分析。考慮前後文發聲變異的影響，自動語音辨識採用右文相關的次音節模型，利用統計式隱藏式馬可夫模型建立語音辨識模組。並把中文分成母音和子音，以次音節為單位建構出可能的辨識音節網路。共定義了 151 個辨識聲學模型，包括 112 個子音和 38 個母音，以及 1 個靜音。

語音及唇形合成

語音合成是把輸入的文字轉換成語音輸出。應用自動語音合成技術，能夠隨機地把給使用者的回應及服務內容朗讀出來，使魔鏡更親近使用者。

語音合成技術是先把輸入文字做語言學預處理，這是對文章段落進行特殊符號、斷詞及斷句的分析，再根據文章中的標點、段落等做文本的劃分，並把它轉換成相應的讀音。接著進行文本分析，就是針對文

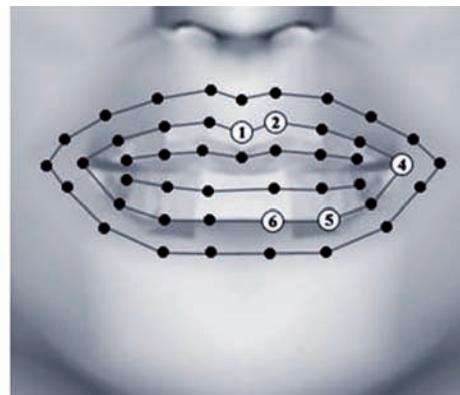
強與音節間的停頓長短，以便正確地表達語意。最後透過電腦音效卡、喇叭等輸出裝置，模擬真人智慧發音。如此，魔鏡可以把網路氣象、新聞等資訊自動朗讀出來。

同時，為使魔鏡精靈模擬真人發音的唇形，首先經由一組事先拍攝的唇形變化影片擷取 62 個唇形特徵點變化參數。另外，根據定義的 40 個聲學模型，分別拍攝其發音的唇形影帶。接著利用光學流動態偵測的方式，自動偵測唇形周圍幾個特徵點的變化。

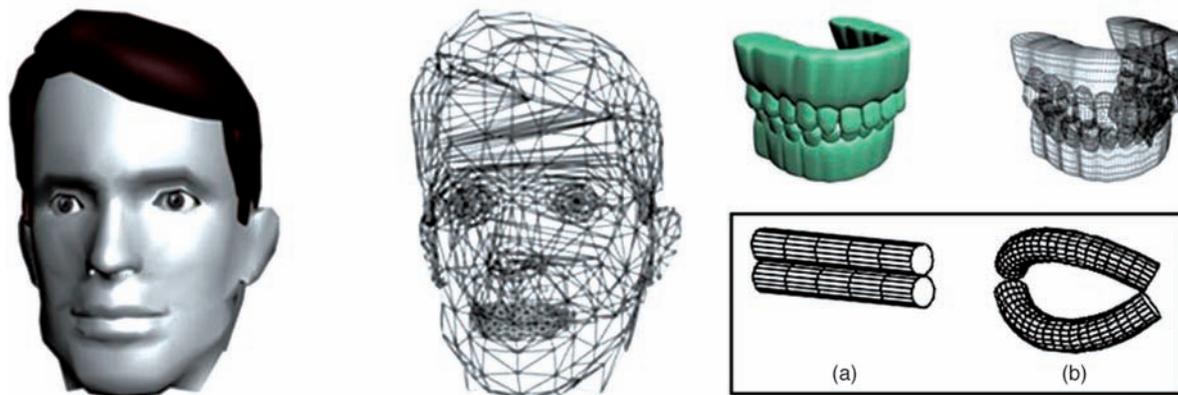
擷取出唇形特徵點在三個座標軸中的位移之後，必須先在 3D 模型中定義出這 62 個特徵點的位置。其餘網格中的點則由鄰近的控制點來控制，位移量是鄰近控制點位移量乘以個別權重的總和，而控制點的權重與控制點到網格點距離的平方成反比。

影像處理

魔鏡可以透過內建攝影機抓取使用者的臉部影像，在對話中進行影像追蹤，使得 3D 虛擬精靈在對話中能注視著使用者，讓人機互動更為真實。在三維空間中，要在不固定的位置與距離中同時做到人臉辨識



利用不同角度的兩個鏡子拍攝實際的唇形動作，偵測唇形特徵點，並擷取 3D 虛擬精靈的唇形特徵參數。



3D 虛擬精靈的人頭、嘴及唇形模型。



原始 (R,G, B)

灰階 (Y, Y, Y)

灰階 (R, R, R)

灰階 (G, G, G)

灰階 (B, B, B)

把色彩轉換後的Y灰階影像與RGB灰階影像進行比較，可以發覺G灰階影像非常近似Y灰階影像，R與B灰階影像則顯得昏暗與變形。因此，單純採用G灰階影像做為發展人臉偵測與特徵點擷取的主要參考資訊。

與人機介面是一件困難的事，且希望前端的人機介面所計算的參數也能夠應用在後端的辨識系統，技術的開發可應用於戶外與戶內等不同背景與光線的情況，並有別於過去採用的邊緣濾波器法來獲得人臉的特徵點。

在人臉偵測方面，使用膚色來偵測人臉，容易遭受光源位置與亮度的影響，而導致較高的錯誤率與較低的精確率。許多的研究成果顯示，單純地採用灰階影像中的資訊來從事類神經網路、面板匹配等方法，都比膚色判斷佳，但是利用膚色的資訊來判斷，的確有助於加快運算速度以及有效輔助人臉偵測。把色彩轉換後的Y灰階影像與RGB灰階影像進行比較，可以發覺G灰階影像非常近似Y灰階影像，R與B灰階影像則顯得昏暗與變形。因此，單純採用G灰階影像做為發展人臉偵測與特徵點擷取的主要參考資訊。

在人臉的偵測與識別時，穩定性最高的可分析資料能夠得到最佳的偵測效果，也就是說這些資訊較不會隨著亮度、人臉形狀與複雜背景的改變而改變。在灰階影像上有哪些人臉五官是較不受以上現象干擾，

可做為人臉偵測的優先判斷準則呢？相較於其他部位，在臉頰的偵測上，臉頰膚色的穩定性與均勻度的表現較佳。

互動式多媒體鏡面系統

影像與語音處理各有本身的限制，尤其是在辨識功能上，例如語音辨識會受限於吵雜的環境所導致的雜訊干擾，影像辨識則受限於影像解析度、複雜的背景與亮度，使得參考資訊不足。結合影音處理技術並應用在數位生活中，只需簡單低成本的設備，非常符合一般大眾的需求。

魔鏡－互動式多媒體鏡面系統的技術是非常廣泛且龐大的，如何整合相關的技術並達到人性化、最佳化、低成本的目標，是未來致力研究的方向。 □

黃建霖

成功大學資訊工程學系

丁俊仁 林基焜 楊家輝

成功大學電機工程學系

吳宗憲

成功大學資訊工程學系