

本功能使用Google Translate翻譯軟體，其翻譯結果僅供參考，本局不保證其翻譯結果之準確、完整或適用於特  
目的。

This translation is machine-generated by using Google Translate. The translation is for reference  
TIPO does not guarantee the translation is accurate, complete, reliable or fit for specific  
purposes.

DTD版本：1.0.0

## 發明專利說明書

※申請案號：093115864

※IPC分類：

### 一、發明名稱：

即時互動影音的方法與系統

METHOD AND SYSTEM FOR REAL-TIME INTERACTIVE VIDEO

### 二、中文發明摘要：

一種即時互動影音的方法與系統，包含：具有畫面之顯示裝置、具有至少一處理器、記憶體及程式之計算機器及攝影裝置。其中程式提供媒體素材及特效指令描述。當攝影裝置接收現場影像且與特效指令描述整合播放時，媒體素材能即時顯示於畫面中並與畫面中的現場影像互動。

### 三、英文發明摘要：

A method and system for real-time interactive video includes a display with a frame, a live person, a computed machine with a processor, memory, a program and a capture device. The program provides media content and effect track script. The capture device receives the live person image and combines the live person image and the effect track script. The program synthesizes the media content and the output of combination with the live person image and the effect track script, and that displays the real-time result on the frame.

### 四、指定代表圖：

(一)本案指定代表圖為：第三B圖

(二)本代表圖之元件符號簡單說明：

401 . . . 現場人員影像

402 . . . 虛擬人物

500 . . . 畫面

501 . . . 臉紅效果

502 . . . 心花怒放效果

503 . . . 一對耳朵

### 五、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

### 六、發明說明：

#### 【發明所屬之技術領域】

[0001] 本發明係有關於動態影音製作的方法與系統，特別是關於一種即時互動式動態影音製作的方法與系統。

#### 【先前技術】

[0002] 由於數位相機、網路視訊及照相機等影像裝置的平價化及普及化，家用電腦與消費性電子產品的結合已是一股無法抵擋的趨勢。然而目前影音多媒體的應用，多限於靜態圖像為主，其重點通常在於相片的拍攝儲存及檔案管理，搭載基本的影像處理與簡單影像合成功能。至於動態影音設備的應用則以單純的錄影、轉檔與播放為主流，有時搭配網路進行即時影音傳訊，而缺乏對多媒體內容的加值創作或改造。另外，有些遊戲軟體雖嘗試整合使用者之肢體動作於互動式的遊戲之中，但囿於運動模式理解層次，在遊戲腳本的設計上有重大的限制，也因此遊戲的內容變化有所限制。

[0003] 此外，電視內容常可看見之特效，因其製作時所需的軟體硬體成本與專業知識過高，因此為一個門檻很高的專業領域。再者，演員必須與空氣對演，憑空想像，對於演員是一大考驗，且後製作上亦不容易。

#### 【發明內容】

[0004] 鑒於上述之發明背景中，一般製作數位內容過於複雜，於此提供一種即時互動影音的方法與系統，提供一輕鬆自然的人機介面，讓一般使用者創作平價而豐富的數位內容。

[0005] 再者，使用互動特效軌(Interactive Effect Track)之概念，在原有的影片中，如：影音軌(video track)、音軌(audio track)之外，即時增添特效之元素；與一般影片特效不同之處，在於本發明所規劃之特效乃即時產生，且其套用對象並非事先擇定，而會伴隨互動而有不同的變化。

[0006] 本發明在提供一種即時互動影音的方法與系統，包含：具有畫面之顯示裝置、現場人員、具有至少一處理器、記憶體及程式之計算機器及一攝影裝置。其中程式提供媒體素材及特效指令描述。攝影裝置接收現場人員影像且與特效指令描述整合。最後，合成媒體素材及整合現場人員影像與特效指令描述之輸出，並即時顯示於畫面。

## 【實施方式】

- [0007] 本發明用示意圖詳細描述如下，在詳述本發明實施例時，表示一種即時互動影音的方法與系統會不依一般比例作局部放大以利說明，然不應以此作為有限定的認知。
- [0008] 一種即時互動影音的方法與系統，包含：具有畫面之顯示裝置、具有至少一處理器、記憶體及程式之計算機器及攝影裝置。其中計算機器中的程式提供：媒體素材及特效指令描述。而當攝影裝置在接收現場影像且與特效指令描述整合播放時，媒體素材能即時顯示於畫面中，而媒體素材可包含一虛擬人物，並即時與畫面中的現場影像互動。
- [0009] 參照第一圖，在一實施例中，提供一具有處理器及記憶體之機器，如：個人電腦、數位機上盒(set-top box)或遊戲機平台(game console)甚至手機等，此處為一電腦主機100。一顯示裝置(displayer)，如：陰極射線管螢幕、液晶顯示螢幕或電漿螢幕等，在此實施例中為液晶顯示螢幕101；以及一攝影裝置(capture device)，在此實施例中為網路攝影機(web-cam)102。這裏要說明的是，此實施例中，電腦主機100、液晶顯示螢幕101及網路攝影機102以有線方式或無線方式相互連接。當然，在此並不受限，主機與顯示器結合，如筆記型電腦或平板電腦，再配上攝影裝置也可應用於此實施例。
- [0010] 接著，一現場實況錄影(live recording)如第一圖中，網路攝影機102對著一現場人員104(live person)。網路攝影機102擷取現場人員104的影像並顯示於液晶顯示螢幕101的畫面103中。在畫面103中顯示出現場人員影像105，且現場人員影像105為即時顯示出之仍立於網路攝影機102前(入鏡)的現場人員104。於一實施例中，在一預選模式中，產生一虛擬人物106與現場人員影像105互動。這裏要說明的是，現場人員104是即時顯現於畫面103而成為現場人員影像105。於此之即時(real time)係指現場人員104的動作與現場人員影像105同步。再者，現場人員104所在的場景及虛擬人物106與現場人員影像105互動方式並不事先設定，而由使用者透過選單或類似介面擇定。預選模式可為程式撰寫好之應用程式，儲存於一記憶體，如電腦主機100中之記憶體。詳細說明如下。
- [0011] 參閱第二圖所示為一實施例中之檔案架構示意圖。預先選擇之模式由主體內容與特效描述檔案所組成，在一實施例中，可先擬定媒體素材201(media content)與腳本以產生多媒體影音內容，例如流行音樂、懷念老歌或經典樂曲等。再者，設計一套相對應的預設互動效果之特效指令描述202(Effect Track Script)，包含時間參數、相對空間參數、特效種類、特效套用對象等基本資訊，並以特定語言描述，存成一指令(Script)檔案。其中使用者可依性別、年齡等因素設計不一樣的主題(theme)而搭配不同效果的特效。即同一主體內容而言，可搭載多項特效指令，舉例來說，播放流行音樂時，相對應的特效描述可載入一虛擬人物，其播放時資料整合之方式，首先使用者下載媒體素材201與特效指令描述202。接著，擷取現場人員影像203搭配影像裝置即時擷取影片，如第一圖中擷取現場人員影像105後，與特效指令描述202串流整合，最後合成動態影音204將串流後之即時擷取影片及特效指令描述202與媒體素材201合成，如此便顯現現場人員融入虛擬世界之中的效果。
- [0012] 如第三A至三B圖所示為一實際擷取現場人員與虛擬世界結合即時播放的示意圖，顯示裝置擷取一畫面，其係從攝影裝置(未顯示)攝影一現場人員(live person)並即時(real-time)顯示於顯示裝置之畫面400，其中存在一現場人員影像401(live person image)。當執行本實施例之可讀程式時，預先選擇之模式可產生一虛擬像，如：人像、神像、卡通人物、妖魔鬼怪等，例如產生一虛擬人物402。
- [0013] 此時，虛擬人物402會與現場人員影像401互動，並即時顯示於畫面400，如第三B圖所示，虛擬人物402可以有許多動作與特效，而現場人員影像401也可左右移動進行小幅度的運動。如實施例中，虛擬人物402的動作為爬上現場人員影像401之肩膀，並親吻現場人員影像401之臉頰。此時回應虛擬人物402的動作，現場人員影像401便產生臉紅效果501，與心花怒放效果502。另一例子是，虛擬人物402也可對現場人員影像401施行魔法，此時回應虛擬人物402的動作，現場人員影像401頭上長一對耳朵503，且當現場人員影像401的頭部有小幅度的擺動時，耳朵503亦會隨之擺動。即虛擬人物402、現場人員影像401與各種效果是即時且互動的。如實施例中，虛擬人物402對現場人員影像401施行魔法，則現場人員影像401頭上長一對耳朵503。此時一對耳朵503是跟著現場人員影像401互動的。也就是說，於畫面500中，現場人員影像401不論移到何處，一對耳朵503永遠在現場人員影像401頭上。於此要說明的是，技術上可先使用辨認(recognition)技術，先確認現場人員影像401的頭髮位置，接著再使用追蹤(tracking)技術追蹤頭部移動位置，再將特效之一對耳朵503加於頭髮上，如此反覆的辨認及追蹤，於畫面上，便產生了人與虛擬物即時的互動效果。
- [0014] 對現場人員而言，現場人員影像可分為半身模式與全身模式，半身模式為現場人員只有頭及肩部位顯示於畫面，全身模式在畫面中為身體部分佔約全身十分之七。於此要說明的是，構成互動式數位內容，即時性與準確性是難以兼得的兩項目標，本發明之實施例中可依照應用型態的不同作適當的處置與調配。舉例來說，應用於動態影音大頭貼時，以臉部特徵偵測與正確定位為主要考量。動作模式以全域動作之簡易參數估算為主。全身模式時則以區域運動之追蹤與組態之辨識為互動模組的重心。
- [0015] 虛擬像與現場人員的互動使用方法，如特徵偵測、特徵追蹤及姿勢分析與辨識等分析虛擬像與現場人員的動作。特徵偵測是依應用目標的性質，分別考慮低階(特徵點)與高階特徵(臉部特徵如眼睛、嘴巴)之擷取。而針對特徵的匹配方式尚有隱性(Implicit)與顯性(Explicit)法則之分。顯性特徵匹配法尋求特徵之間的一對一對應關係(one to one correspondence)；隱性匹配法則以參數或轉換(transformation)等方式代表前後畫格中特徵之間的關係。如：顯性法則及低階特徵可為特徵點匹配(肢體追蹤)、顯性法則及高階特徵可為表情分析、隱性法則及低階特徵可為密集光流分析(Dense optical flow)以及隱性法則及高階特徵可為臉部器官偵測與定位。
- [0016] 特徵偵測中，使用下列方法，作有效率且精準的人臉偵測與器官定位。初始偵測，一實施例中，以灰階影像中水平邊緣之密度強弱初估眼睛與嘴的可能位置，如第四圖所示為水平邊緣密度計算之初步選定連續圖。候選區域601即為所選定之眼睛與嘴的可能位置。其次，在眾多候選區域601中，利用器官相對位置與比例關係做進一步篩選。最後，再利用眼球搜尋做位置確認。在一實施例中，也可將膚色作為輔助判斷依據。器官定位，一實施例中，如鼻子、眉毛及耳朵等，採用比例關係做位置估計。而人臉的外框則以橢圓方程式表示。在一實施例中，如為全身操作模式下，可透過膚色模型搭配髮特徵偵測器(Hair-Like Feature Detector)作快速偵測，至於人體其他部分我們以較低階但經「群組化」的特徵點描述之。
- [0017] 特徵追蹤(Feature Tracking)，對於半身操作模式的特徵追蹤之一實施例中，將著重於臉部器官的持續定位與整體區塊的運動參數估算。至於全身操作模式之一實施例中，將以群組化的圖形匹配(Graph Matching)方式作

特徵的比對與追蹤，並視計算資源的多寡動態調整特徵點的數目。要說明的是，本實施例中，現場人員的臉部應遷就擷取影像裝置的鏡頭，而非由擷取影像裝置的鏡頭去追蹤現場人員的臉部，如此可無須考慮姿勢的估算(pose estimation)。

- [0018] 姿勢分析與辨識(Gesture Analysis and Recognition)，靜止狀態下物件組態(Configuration)的判別之一實施例中，是可使用形狀比對(Shape Matching)，其相關技術，如Shape Context，而演算法也可為Elastic Matching演算法，並配合多重解析度之概念，以容忍小幅度的變形(Deformation)以及遮蔽(Occlusion)效應。關於連續動作之分析與辨識之一實施例中，利用階層式光流追蹤的方式(Pyramidal Optical Flow)，先計算出人體的移動方向與速率，在使用時間序列法之一實施例中，可為Hidden Markov Model(HMM)或Recurrent Neural Network(RNN)等，以分析該動作所代表的意義。
- [0019] 參閱第五圖，其為本發明軟體運作流程圖之一實施例。首先觸發應用程式701，偵測硬體751，警告訊息731，終止應用程式704，及問題訊息732為程式認證硬體需求之步驟。當偵測硬體751發現問題時則產生警告訊息731，反之則產生問題訊息732。警告訊息731為提醒使用者在硬體偵測時所需之硬體設備未安裝或無法運作，例如，攝影鏡頭未安裝或攝影鏡頭安裝不完全等訊息。問題訊息732為提示使用者先離開鏡頭，以便接下來的取景步驟，接下來為前置處理，收集背景資料706存入內部儲存背景資料707中，接著產生問題訊息733，其目的為重新邀請使用者進入鏡頭。例如，一歡迎畫面邀請使用者進入鏡頭且其影像出現於顯示畫面。
- [0020] 辨認709在此可辨認臉及整個肢體。追蹤動作710在此可偵測臉及整個肢體動作。另外媒體資料761，其可包含延伸檔案類型如AVI或MPEG格式。在一實施例中，媒體資料761可為壓縮檔案，如：DLL檔。接著載入媒體資料711及媒體資料解碼713。辨認709、追蹤動作710以及內部儲存背景資料707與接下來的步驟配合便可產生動態合成影音。
- [0021] 合成攝影機影像及媒體資料714及動作再追蹤715後，顯示合成媒體資料716。動作再追蹤715為再一次偵測背景及影像之改變。接著判斷是否載入特效752，是，則進入步驟載入嵌入特效718。載入嵌入特效718，在一實施例中，其特效等級可為"CEffect"。接著，是否儲存合成媒體資料753，是，則儲存合成媒體資料720。時間是否結束754。是，則進入再處理儲存合成媒體資料722，在一實施例中，可為JPEG檔案格式，或可為CStyle等級。最後顯示再處理儲存合成媒體資料723及終止應用程式724。
- [0022] 這裏要說明的是，合成攝影機影像及媒體資料714經動作再追蹤715後，便可顯示合成媒體資料716顯示於畫面上。接著載入嵌入特效718，經儲存合成媒體資料720後，便進入迴圈至合成攝影機影像及媒體資料714，如此便產生即時之效果。對照第三A及第三B圖，虛擬人物402經動作再追蹤715後便可知現場人員影像401之肩膀及臉頰位置。而當特效臉紅效果501經儲存合成媒體資料720及動作再追蹤715後，便可即時見到如第三B圖之臉紅效果501。且在這之中因動作再追蹤715，不論臉頰移動至何處，臉紅效果501都會產生在正確的位置上。
- [0023] 以上所述僅說明本發明一軟體運作流程圖之一實施例。而本發明更可透過個人電腦(PC or laptop)、數位機上盒(set-top box)或遊戲機平台(game console)甚至手機等上執行。而在應用上，兩使用者更可互相對玩。兩使用者可透過網路，如internet或intranet連結，並為對方或己方選擇虛擬人物，在其中一端下指令，遙控另一端的虛擬人物，並做出各種不同的視覺特效，結果可顯示在對方及自己的顯示器上。
- [0024] 根據上述，本發明之一實施例中，兼顧應用軟體的互動性與合成效果的逼真度，將特效模組與互動模組的設計一併考量，並結合成一個封包(package)，如此可於媒體內容編排時就先行處理完畢，使系統資源得以充分利用於互動時的逼真呈現。
- [0025] 以上所述僅為本發明之較佳實施例而已，並非用以限定本發明之申請專利範圍；凡其它未脫離本發明所揭示之精神下所完成之等效改變或修飾，均應包含在下述之申請專利範圍中。

#### 【圖式簡單說明】

- [0069] 第一圖為根據本發明之一實施例之架構示意圖。
- [0070] 第二圖所示為一實施例中之檔案架構示意圖。
- [0071] 第三A至三B圖所示為一實際擷取現場人員與虛擬世界結合即時播放的示意圖。
- [0072] 第四圖顯示一符合本發明之一實施例應用水平邊緣密度計算之初步選定連續圖。
- [0073] 第五圖顯示本發明軟體運作流程圖之一實施例

#### 【主要元件符號說明】

- [0026] 100 . . . 電腦主機
- [0027] 101 . . . 液晶顯示螢幕
- [0028] 102 . . . 網路攝影機
- [0029] 103 . . . 畫面
- [0030] 104 . . . 現場人員
- [0031] 105 . . . 現場人員影像
- [0032] 106 . . . 虛擬人物
- [0033] 201 . . . 媒體素材
- [0034] 202 . . . 特效指令描述
- [0035] 203 . . . 擷取現場人員影像
- [0036] 204 . . . 動態合成影音
- [0037] 400 . . . 畫面
- [0038] 401 . . . 現場人員影像
- [0039] 402 . . . 虛擬人物
- [0040] 500 . . . 畫面
- [0041] 501 . . . 臉紅效果
- [0042] 502 . . . 心花怒放效果

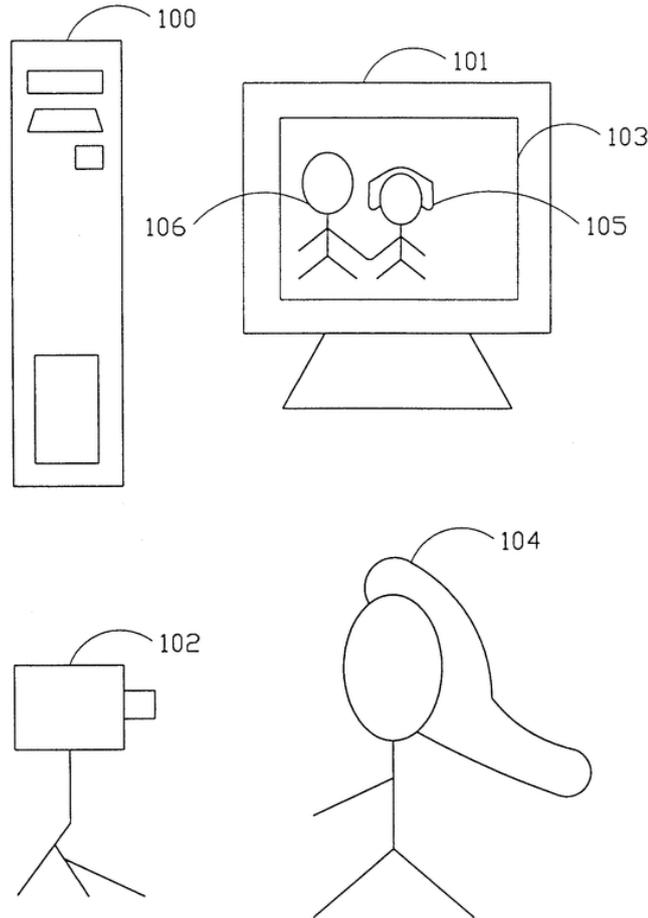
- [0043] 503 . . . 一對耳朵
- [0044] 601 . . . 候選區域
- [0045] 701 . . . 觸發應用程式
- [0046] 751 . . . 偵測硬體
- [0047] 731 . . . 警告訊息
- [0048] 704 . . . 終止應用程式
- [0049] 732 . . . 問題訊息
- [0050] 706 . . . 收集背景資料
- [0051] 707 . . . 內部儲存背景資料
- [0052] 733 . . . 問題訊息
- [0053] 709 . . . 辨認
- [0054] 710 . . . 追蹤動作
- [0055] 711 . . . 載入媒體資料
- [0056] 761 . . . 媒體資料
- [0057] 713 . . . 媒體資料解碼
- [0058] 714 . . . 合成攝影機影像及媒體資料
- [0059] 715 . . . 動作再追蹤
- [0060] 716 . . . 顯示合成媒體資料
- [0061] 752 . . . 是否載入特效
- [0062] 718 . . . 載入嵌入特效
- [0063] 753 . . . 是否儲存合成媒體資料
- [0064] 720 . . . 儲存合成媒體資料
- [0065] 754 . . . 時間是否結束
- [0066] 722 . . . 再處理儲存合成媒體資料
- [0067] 723 . . . 顯示再處理儲存合成媒體資料
- [0068] 724 . . . 終止應用程式

## 七、申請專利範圍：

1. 一種即時互動影音的方法，包含：提供一畫面(frame)；即時(real time)接收一影像(video)顯示於該畫面；產生一物件並顯示於該畫面；及進行互動於該物件與該影像之間。
2. 如申請專利範圍第1項所述之即時互動影音的方法，其中接收該影像之方法為使用一網路攝影機(web cam)接收面對該網路攝影機的影像。
3. 如申請專利範圍第1項所述之即時互動影音的方法，其中該產生步驟更包含一預選模式以產生對應之物件。
4. 如申請專利範圍第1項所述之即時互動影音的方法，其中該進行互動步驟包含辨認該影像的位置。
5. 如申請專利範圍第1項所述之即時互動影音的方法，其中該進行互動步驟包含追蹤該影像之一變動。
6. 如申請專利範圍第1項所述之即時互動影音的方法，其中該進行互動步驟包含產生一特效描述於該影像上。
7. 如申請專利範圍第6項所述之即時互動影音的方法，其中該特效描述係選自一以描述語言描述的特效指令集。
8. 如申請專利範圍第1項所述之即時互動影音的方法，其中該物件係選自一媒體素材(media content)。
9. 如申請專利範圍第1項所述之即時互動影音的方法，其中該提供步驟包含一影像提供者存在於一攝像裝置前且該攝像裝置與該畫面有電性上的連接。
10. 如申請專利範圍第1項所述之即時互動影音的方法，其中產生步驟包含該影像依循該影像之一特徵追蹤。
11. 如申請專利範圍第1項所述之即時互動影音的方法，其中產生步驟包含對該影像依循一姿勢分析與辨識。
12. 一種儲存裝置，儲存可由媒體處理裝置判讀之複數個程式，其中該媒體處理裝置根據該複數個程式執行的步驟包含：輸入一資料，其中該資料包含一背景資料及一即時影像；辨認該資料；追蹤該資料改變之部分；提供一媒體素材；合成(synthesizing)該媒體素材及該資料；及顯示合成該媒體素材及該資料。
13. 如申請專利範圍第12項所述之儲存裝置，儲存可由媒體處理裝置判讀之複數個程式，其中提供一媒體素材步驟更包含：載入該媒體素材；及解碼該媒體素材。
14. 如申請專利範圍第12項所述之儲存裝置，儲存可由媒體處理裝置判讀之複數個程式，其中合成(synthesizing)該媒體素材及該資料步驟更包含：再追蹤該資料改變之部分。

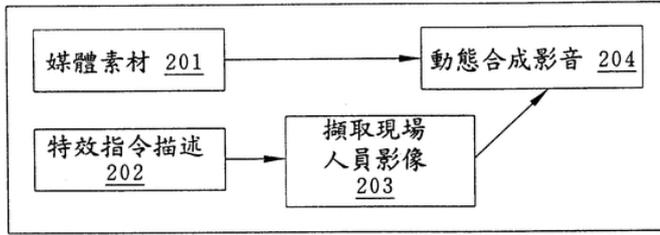
15. 如申請專利範圍第12項所述之儲存裝置，儲存可由媒體處理裝置判讀之複數個程式，其中更包含：載入特效；再處理合成該媒體素材、該資料及特效；及顯示合成該媒體素材、該資料及特效。
16. 如申請專利範圍第15項所述之儲存裝置，儲存可由媒體處理裝置判讀之複數個程式，其中步驟載入特效更包含嵌入特效於該背景資料。
17. 一種即時互動影音的系統，包含：一顯示裝置，具有一畫面；一計算器具有至少一處理器、一記憶體及複數個可讀程式，其中該複數個可讀程式具有一媒體素材及一特效指令描述；及一攝影裝置接收一影像，經該特效指令描述處理並與該媒體素材合成，即時顯示該媒體素材與該影像的互動於該畫面。
18. 如申請專利範圍第17項所述之即時互動影音的系統，其中該顯示裝置為一液晶螢幕。
19. 如申請專利範圍第17項所述之即時互動影音的系統，其中該計算器為一電腦。

#### 八、圖式：



第一圖

第一圖



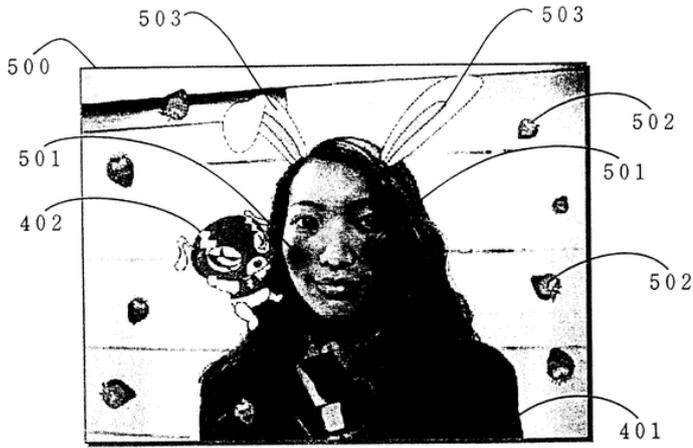
第二圖

第二圖



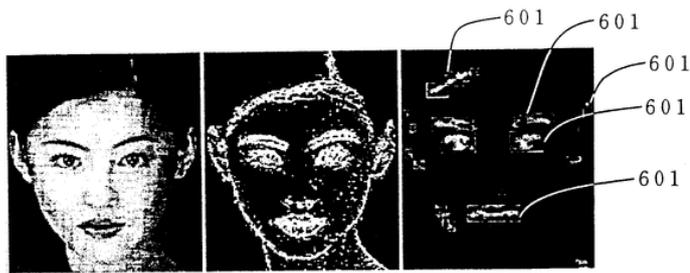
第三A圖

第三A圖



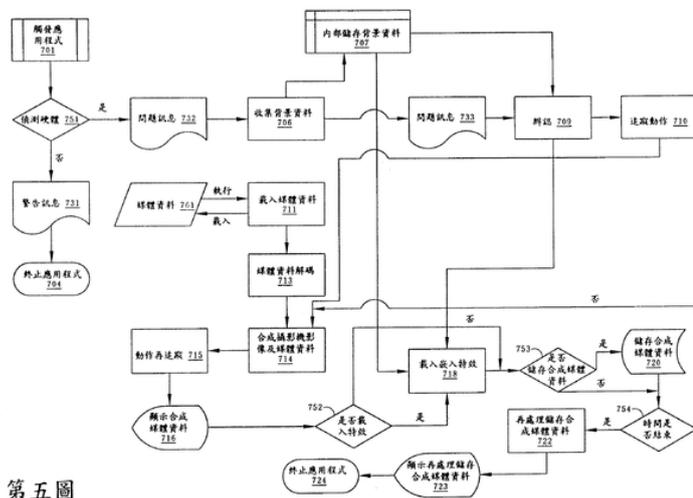
第三B圖

第三B圖



第四圖

第四圖



第五圖

第五圖