



(19) 中華民國智慧財產局

(12) 發明說明書公告本

(11) 證書號數：TW I682368 B

(45) 公告日：中華民國 109 (2020) 年 01 月 11 日

(21) 申請案號：107122943

(22) 申請日：中華民國 107 (2018) 年 07 月 03 日

(51) Int. Cl. : **G08B19/00 (2006.01)** **G06N20/00 (2019.01)**

(71) 申請人：緯創資通股份有限公司 (中華民國) WISTRON CORP. (TW)

新北市汐止區新台五路 1 段 88 號 21 樓

(72) 發明人：郭方文 KUO, FANG WEN (TW) ; 陳志明 CHEN, CHIH MING (TW)

(74) 代理人：洪澄文

(56) 參考文獻：

TW M482134

TW M547727

CN 107767617A

US 2013/0061311A1

US 2016/0364927A1

WO 2016/088401A1

WO 2016/109039A1

審查人員：李志偉

申請專利範圍項數：18 項 圖式數：11 共 97 頁

(54) 名稱

利用多維度感測器資料之監控系統及監控方法

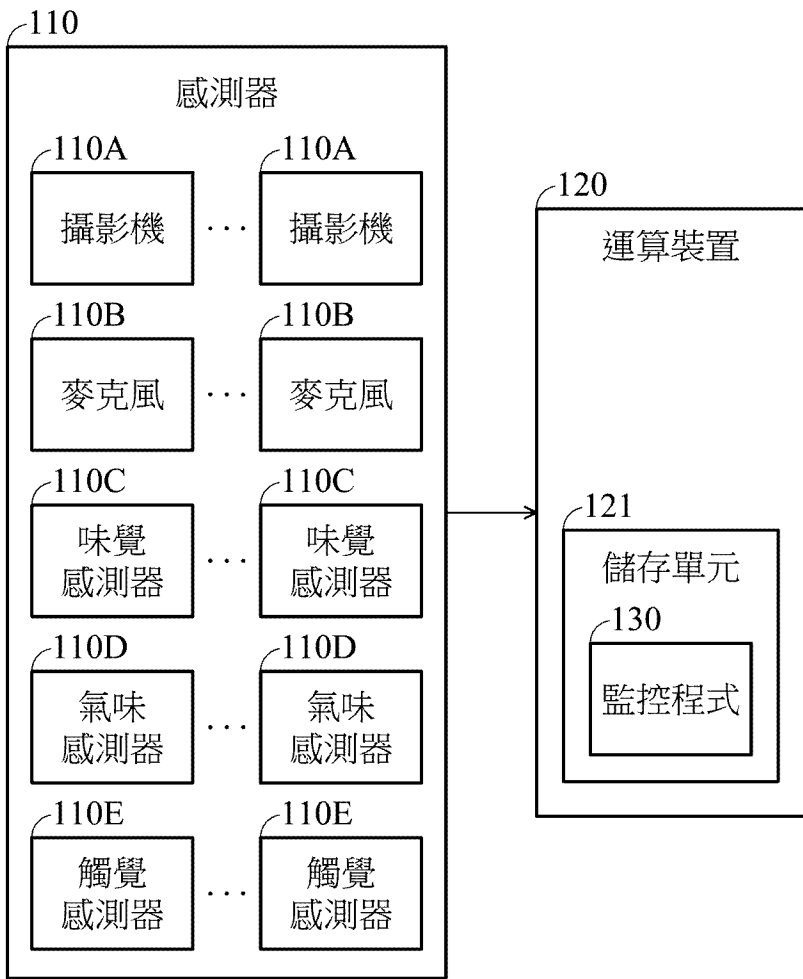
(57) 摘要

一種利用多維度感測器資料的監控方法，用於一監控系統。該監控系統包括設置於一場景中的複數個感測器，且該等感測器係分為複數個類型。該監控方法包括：利用該等感測器偵測該場景以取得各類型的一感測資料；分別對各類型的該感測資料進行一本地物件處理以產生一本地物件特徵資訊；依據該本地物件特徵資訊進行一全局物件處理以產生一全局物件特徵資訊；以及對該全局物件特徵資訊進行一全局物件辨識以產生一全局辨識結果。

A surveillance method using multi-dimensional sensor data for use in a surveillance system is provided. The surveillance system includes a plurality of sensors disposed in a scene, and the sensors are classified into a plurality of types. The surveillance method includes the steps of: utilizing the sensors to detect the scene to obtain sensor data of each type; performing a local object process on the sensor data of each type to generate local object feature information; performing a global object process according to the local object feature information to generate a global object feature information; and performing a global object recognition on the global object feature information to generate a global recognition result.

指定代表圖：

100



符號簡單說明：

- 100 . . . 監控系統
- 110 . . . 感測器
- 120 . . . 運算裝置
- 121 . . . 儲存單元
- 130 . . . 監控程式
- 110A . . . 攝影機
- 110B . . . 麥克風
- 110C . . . 味覺感測器
- 110D . . . 氣味感測器
- 110E . . . 觸覺感測器

第 1 圖

發明專利說明書

【發明名稱】 利用多維度感測器資料之監控系統及監控方法

Surveillance System and Surveillance Method Using
Multi-Dimensional Sensor Data

【技術領域】

【0001】 本發明係有關於監控系統，特別是有關於一種利用多維度感測器資料之監控系統及監控方法。

【先前技術】

【0002】 為了維護私人與公眾的身家財產與交通行車安全，預防及打擊罪犯等需求，視訊監視系統或視訊攝影機已廣泛設置於居家、私人區域、公共場所、及交通要道，用以即時監控或錄影存證。然而，傳統的視訊監視系統只能持續地錄製視訊，且因大量裝設，長時間下來產生了極為龐大的視訊資料。當特殊事件發生的當下，不一定能被及時發現與處置，且得再耗費大量的人力與時間去調閱及觀看錄影內容。

【0003】 然而，在現實環境中，單靠視訊攝影機並無法收集到完整的資訊。例如：當火災剛發生或是油氣洩漏剛發生的時候，在隱密處尚未被察覺，災害尚未擴大前，透過偵測空氣中飄散異常氣味，可預先探知可能將有災害發生。因此，單一類型的感測器對於安全監控與維護其實是有所限制的。

【發明內容】

【0004】 本發明實施例係提供一種利用多維度感測器資料的監控方法，用於一監控系統，該監控系統包括設置於一場景中的複數個感測器，且該等感測器係分為複數個類型，該監控

方法包括：利用該等感測器偵測該場景以取得各類型的一感測資料；分別對各類型的該感測資料進行一本地物件處理以產生一本地物件特徵資訊；依據該本地物件特徵資訊進行一全局物件處理以產生一全局物件特徵資訊；以及對該全局物件特徵資訊進行一全局物件辨識以產生一全局辨識結果。

【0005】 本發明實施例更提供一種監控系統，包括：複數個感測器，其中該等感測器係分為複數個類型，且用以偵測一場景以取得各類型的一感測資料；以及一運算裝置，用以分別對各類型的該感測資料進行一本地物件處理以產生一本地物件特徵資訊，其中該運算裝置更依據該本地物件特徵資訊進行一全局物件處理以產生一全局物件特徵資訊，並對該全局物件特徵資訊進行一全局物件辨識以產生一全局辨識結果。

【圖式簡單說明】

【0006】

第1圖係顯示依據本發明一實施例中之監控系統的功能方塊圖。

第2圖係繪示依據本發明一實施例中之監控程式的功能方塊圖。

第3A及3B圖係顯示依據本發明一實施例中用於視頻物件之本地物件對應處理、本地細緻特徵融合處理、與本地物件身份辨識的流程圖。

第4A及4B圖係顯示依據本發明一實施例中用於音頻物件之本地物件對應處理、本地細緻特徵融合處理、與本地物件身份辨識的流程圖。

第5圖係顯示依據本發明一實施例中之全局物件對應處理及建立全局細緻特徵集之流程圖。

第6A圖係顯示依據本發明一實施例中利用多個攝影機對一場景擷取視頻資料的示意圖。

第6B圖係顯示依據本發明一實施例中利用多個麥克風對一場景擷取音訊資料的示意圖。

第7A圖係顯示依據本發明一實施例之視頻幀中之不同空間分割方式的示意圖。

第7B圖係顯示依據本發明一實施例之音頻段中之不同時間分割方式的示意圖。

第8A圖係顯示依據本發明一實施例中之依據上下文分析處理以選擇全局細緻特徵融合之係數的流程圖。

第8B圖係顯示依據本發明第8A圖之實施例中全局上下文分析處理及權重判斷步驟的流程圖。

第8C-1及8C-2圖係顯示依據本發明一實施例中之全局細緻特徵融合及全局物件辨識之流程圖。

第8D圖係顯示依據本發明一實施例中之全局細緻特徵融合與全局物件辨識之資料管線的示意圖。

第8E圖係顯示依據本發明一實施例中之辨識結果回饋與強化之全局回饋的流程圖。

第8F圖係顯示依據本發明一實施例中之辨識結果回饋與強化之本地回饋的流程圖。

第9A-1及9A-2圖係顯示依據本發明一實施例中之監控方法的方塊圖。

第9B-1及9B-2圖係顯示依據第9A-1及9A-2圖之實施例之全局上下文分析處理之詳細方塊圖。

第9C圖係顯示依據本發明一實施例中之利用多維度感測器資料之監控方法的流程圖。

第10圖係顯示依據本發明一實施例中之一場景及監控系統的示意圖。

第11圖係顯示依據本發明一實施例中之利用多維度感測器資料的監控方法之流程圖。

【實施方式】

【0007】 為使本發明之上述目的、特徵和優點能更明顯易懂，下文特舉實施例，並配合所附圖式，作詳細說明如下。

【0008】 第1圖係顯示依據本發明一實施例中之監控系統的功能方塊圖。

【0009】 如第1圖所示，監控系統100係包括複數個感測器110、一或多個運算裝置120。

【0010】 在一實施例中，感測器110係包括多種不同類型的感測器，例如：攝影機110A、麥克風110B、味覺感測器110C、氣味感測器110D、觸覺感測器110E、或其組合，但本發明之實施例並不限定於上述類型或屬性的感測器，且各個類型或屬性的感測器的安裝數量可視實際情況而定。其中上述不同類型的感測器可分別對應至人類的五個感官，例如攝影機110A可對應至眼，並用於擷取視頻影像；麥克風110B可對應至耳，用於接收音訊；味覺感測器110C(例如可為一電子舌)可對應至舌，用以偵測物件的酸甜苦辣鹹味；氣味感測器110D(例如可為一電

I682368

發明摘要**【發明名稱】**利用多維度感測器資料之監控系統及監控方法**Surveillance System and Surveillance Method Using
Multi-Dimensional Sensor Data****【中文】**

一種利用多維度感測器資料的監控方法，用於一監控系統。該監控系統包括設置於一場景中的複數個感測器，且該等感測器係分為複數個類型。該監控方法包括：利用該等感測器偵測該場景以取得各類型的一感測資料；分別對各類型的該感測資料進行一本地物件處理以產生一本地物件特徵資訊；依據該本地物件特徵資訊進行一全局物件處理以產生一全局物件特徵資訊；以及對該全局物件特徵資訊進行一全局物件辨識以產生一全局辨識結果。

【英文】

A surveillance method using multi-dimensional sensor data for use in a surveillance system is provided. The surveillance system includes a plurality of sensors disposed in a scene, and the sensors are classified into a plurality of types. The surveillance method includes the steps of: utilizing the sensors to detect the scene to obtain sensor data of each type; performing a local object process on the sensor data of each type to generate local object feature information; performing a global object process according to the local object feature

information to generate a global object feature information; and performing a global object recognition on the global object feature information to generate a global recognition result.

【代表圖】

【本案指定代表圖】：第（1）圖。

【本代表圖之符號簡單說明】：

100～監控系統；

110～感測器；

120～運算裝置；

121～儲存單元；

130～監控程式；

110A～攝影機；

110B～麥克風；

110C～味覺感測器；

110D～氣味感測器；

110E～觸覺感測器。

【本案若有化學式時，請揭示最能顯示發明特徵的化學式】：

無。

子鼻)可對應至鼻，用於偵測空氣中之味道；觸覺感測器 110E(例如可為一電子皮膚)可對應至身，用於偵測接觸的壓力或溫度等等。各種類型的感測器的感測資料例如可視為一個維度，本發明之實施例中的監控系統 100 可利用多種類型的感測器的感測資料。

【0011】 在一些實施例中，對應至眼睛之攝影機 110A 可包括不同類型的攝影機，例如攝影機 110A 可為擷取彩色影像(例如 RGB 影像)之彩色攝影機，其可擷取場景之彩色影像；攝影機 110A 亦可為一深度攝影機，其可擷取場景之深度影像(例如為一灰階影像)；攝影機 110A 亦可為一紅外線感測器(infrared sensor)，其可偵測場景中之幅射能量，並將所偵測的幅射能量轉換為電訊號，並以不同顏色來顯示出不同溫度的分佈，例如可由一紅外線熱影像表示。為了便於說明，在後述實施例中，攝影機 110A 係以彩色攝影機為例進行說明。

【0012】 運算裝置 120 例如為一或多台個人電腦或伺服器、或是一中央資料處理中心，用以執行可使用多維度感測器資料之一監控程式 130，其中監控程式 130 例如可儲存於運算裝置 120 中之一儲存單元 121。儲存單元 121 為一非揮發性記憶體，例如是硬碟機(hard disk drive)、固態硬碟(solid-state disk)、或唯讀記憶體(read-only memory, ROM)等等，但本發明之實施例並不限於此。運算裝置 120 係執行監控程式 130，從具有不同類型的各感測器 110 接收相應的感測資料，並進行本地物件辨識處理、本地物件特徵抽取及融合處理、特徵資訊回饋及強化處理、及全局辨識處理之功能，其細節將詳述於後。

【0013】在一實施例中，攝影機 110A、麥克風 110B、味覺感測器 110C、氣味感測器 110D、觸覺感測器 110E 係分別以眼、耳、鼻、舌、身的感測模式進行偵測 (Detection) 與辨識 (Recognition) 物件的動作。然而，偵測物件過程及辨識物件的過程中所抽取的物件特徵是不同的。偵測物件過程係使用粗糙特徵 (Rough Feature)，辨識物件過程則是使用細緻特徵 (Detail Feature)。舉例來說，粗糙特徵有方向、距離、概略、結構等，但本發明之實施例並不限於此。細緻特徵亦可針對不同類型的感測器進行分類，例如視頻細緻特徵包括：顏色、紋理及形狀等；音頻細緻特徵包括：音量、音高、音色等；氣味細緻特徵包括：芳香味、腐敗味、醚味、辛辣味、焦臭味、樹脂味等；味道細緻特徵包括：甜、鹹、酸、苦、辣、鮮等；觸感細緻特徵包括：壓力、溫度等，但本發明之實施例並不限於此。

【0014】舉例來說，物件的結構粗糙特徵用於描述物件輪廓，例如：圓柱形的桶子，長方形的立牌，人，汽車，摩托車等。聲音的結構就是聲紋。聲紋可表現出發聲器官的差異，因為人的聲音是多種頻率聲音的集合。每個人的聲紋不同，因每個人的發聲器官（聲帶、口鼻腔）、嘴唇和舌頭的形狀差異。聲音的三要素為音色、音量、及頻率，其中「音色」取決於口鼻腔的容積以及結構。因此根據聲紋的特點，就可以大致掌握發聲者的年齡、性別、臉型，甚至於身高等粗糙的特徵。

【0015】在一實施例中，運算裝置 120 可計算各攝影機 110A 所擷取之視頻影像之顏色直方圖以獲取顏色的分佈資訊，再以概率質量函數 (probability mass function, PMF) 計算就可以知

道概略的顏色特徵。運算裝置 120 可分析麥克風 110B 所擷取之音頻信號以獲得音頻頻譜，再計算音頻信號之頻率的分佈資訊，再以概率密度函數(probability density function, PDF)計算出概略的頻率特徵。

【0016】 化學性的感覺可以分為偵測液態化學物質的味覺(taste)及偵測氣體中物質的嗅覺(smell)。舉例來說，運算裝置 120 可由氣味感測器 110D 所擷取到之感測資料以獲取七種味道的感知分子(例如：樟腦味、麝香味、花卉味、薄荷味、乙醚味、辛辣味和腐腥味)的分佈資訊，再以概率質量函數計算就可以知道概略的嗅覺特徵。運算裝置 120 亦可由味覺感測器 110C 所擷取到之感測資料以獲取六種味道(例如：甜、鹹、酸、苦、辣、鮮)的分佈資訊，再以概率密度函數計算出概略的味覺特徵。

【0017】 觸覺係指皮膚受到機械刺激時所引起的感覺。觸點在皮膚表面的分佈密度以及大腦皮層對應的感受區域面積與該部位對觸覺的敏感程度呈正相關。舉例來說，運算裝置 120 可由觸覺感測器 110E 所擷取到之感測資料獲取三種物理性質(例如：種類、強度、大小)的分佈資訊，再以概率質量函數計算就可以知道概略的觸覺特徵。

【0018】 詳細而言，在監控系統 100 中，可在場景中安置指向性的遠距化學感受器(例如氣味感測器 110D)、指向性的接觸化學感受器(例如味覺感測器 110C)、及指向性的機械感受器(例如觸覺感測器 110E)，並可利用上述實施例中所揭示的定位技術以計算出在場景中之物件的方向、距離的粗糙特徵，並可概

略地訂出物件所在的空間及移動的向量值。

【0019】 機率密度函數 (PDF) 和機率質量函數 (PMF) 是描述隨機變量 (random variables) 的輸出值，在某個確定的取值點附近的可能性的函數。借由前幾名最高機率的取值，以及這些取值的機率相互間的大小比率，我們可以得到物件概略的特徵。

【0020】 在一實施例中，氣味感測器 110D 或味覺感測器 110C 可安裝於一移動載體上或機動電子警犬身上，如同擁有敏銳嗅覺的緝私犬或者警犬，在一場域內來回搜索探詢，其身上配置定位系統，再透過無線傳輸將探測到的氣味或味道資訊傳送到資料處理中心內，再與其他類型的特徵資料配對運用。

【0021】 在另一實施例中，假設某區域其中心點向東偏北 30 度的扇型區域一公尺內有一位小孩，五公尺外有兩位成人，其中一人穿白色衣服，另一位是黑色衣服。白色成人向北以低於 3.0 公里/小時 (0.83 公尺/秒) 之速度慢速移動，黑色成人向西以高於 12 公里/小時 (3.33 公尺/秒) 之速度快速移動，並發出振幅變化頻率介於 30~150 次/秒之間的尖叫聲。運算裝置 120 可利用以三維攝影機/深度攝影機或指向性麥克風三角定位的技術計算出三個人物相應的方向及距離等等的粗糙特徵，概略地訂出物件所在的空間及移動的向量值。接著，運算裝置 120 可分析在此場景中的視頻影像及音頻信號以得到此三個人物的顏色直方圖及音色等資訊，並藉以得到此三個人物之概略及結構等等的粗糙特徵。

【0022】 需注意的是，各類型之感測資料的粗糙特徵與細

緻特徵的差異在於取樣與取值的細緻度與精確度。粗糙特徵只需要判定資料在一範圍區間內。以顏色來說，粗糙特徵可利用顏色直方圖來粗略判斷顏色，若要取得細緻特徵就需要進行各種影像特徵比對及計算，故所需的計算量極大，但是準確度較高，故細緻特徵可以用於辨識身分，粗糙特徵例如可用於進行簡單分類。以紋理來說，運算裝置 120 也是判定線條圖樣 (pattern)、以及直向線條與橫向線條的數量。依據取樣與取值的細緻度與精確度，資料量越多也表示所取出的特徵越細緻，但也表示計算量大且計算時間也會比較長。

【0023】 以音頻資料為例，一般人的耳朵能聽見的音量範圍在 0~140 分貝 (dB) 之間。分貝是用來表示聲音強度或是音量的單位。40~50 分貝的聲音會干擾睡眠，60~70 分貝的聲音會干擾學習，120 分貝以上的噪音量會導致耳痛，甚至聽力永久喪失。

【0024】 以下為一些音量的例子，例如：樹葉的摩擦聲：20 分貝。郊區的深夜：30 分貝。郊區之住宅區的深夜：40 分貝。安靜的辦公室裡：50 分貝。人平常講話的聲音：60 分貝。車內的聲音或電話的鈴聲：70 分貝。大巴士的行進聲：80 分貝。狗的叫聲：90 分貝。電車通過鐵橋的聲音：100 分貝。汽車的喇叭聲、警笛聲、卡拉 OK 舞廳內的音樂聲：110 分貝。修馬路時打碎機的聲音：120 分貝。噴射機起飛時的引擎聲：130 分貝。

【0025】 人耳是一個非常特別的器官，為了將噪音儀器量測到噪音訊號，轉換成人耳聽到感受到的噪音量，必須對噪音儀器量測到不同頻率域的訊號，進行頻域加權或進行權重計

算。人耳可聽到聲音的頻率範圍約為 20-20K 赫茲 (Hz)，而對於不同聲音強度人耳又有不同的加權曲線 (Weighting Curve)。最常見的加權曲線為 A 加權 (A-Weighting)、C 加權、D 加權、及 G 加權。其中，C 加權 (C-Weighting) 一般用於測定較為吵雜的機械噪音，D 加權 (D-Weighting) 一般用於測定航空噪音，而 G 加權 (G-Weighting) 則用於測定超低頻噪音，大部分皆為低頻振動引起的結構噪音。

【0026】 音色即因發聲體的諧音（泛音）成份比例不同而產生的不同聲音。大自然中的任何聲音皆為複雜的波形，這複雜的波形除了基本頻率的波形之外，還會有一系列的諧振頻率，也就是所謂的「泛音」(Harmonic)，它與基音有一定的「倍音」關係。例如某物體振動之基本頻率為 240Hz，也會發生 480Hz（二次諧波），720Hz（三次諧波）等頻率，每一個物體的倍音組成成份比例都不相同，這種不同物體發生不同的倍音成份的聲音就是音色 (Timbre)。

【0027】 第 2 圖係繪示依據本發明一實施例中之監控程式 130 的功能方塊圖。監控程式 130 例如包括一本地物件辨識模組 131、一特徵融合模組 132、及一全局辨識模組 133。本地物件辨識模組 131 係用以對各類型的該感測資料進行一本地物件處理以產生一本地物件特徵資訊。

【0028】 舉例來說，本地物件辨識模組 131 係包括一本地物件偵測及對應模組 1311、一本地物件特徵抽取及融合模組 1312、及一本地物件辨識模型 1313。本地物件處理例如包括本地物件偵測及對應模組 1311、一本地物件特徵抽取及融合模組

1312、及一本地物件辨識模型 1313 所執行的關於本地物件的各種處理。

【0029】本地物件偵測及對應模組 1311 係分別接收來自攝影機 110A、麥克風 110B、味覺感測器 110C、氣味感測器 110D、及觸覺感測器 110E 之感測資料，並進行相應的感測類型之一本地物件偵測及對應處理以產生一本地物件識別碼清單 (local object ID list, LOID list) 及本地粗糙特徵集 (local rough feature set, LRFS)，其細節將詳述於後。

【0030】本地物件特徵抽取及融合模組 1312 係執行一本地特徵抽取及融合處理，其包括一本地細緻特徵抽取 (local detail feature extraction, LDFE) 處理及一本地細緻特徵融合 (local detail feature fusion, LDFE) 處理。舉例來說，運算裝置 120 係依據本地物件辨識模組 131 所產生的本地物件識別碼清單及本地粗糙特徵集，對各不同類型之感測資料抽取其本地細緻特徵並建立各類型的感測資料相應的本地細緻特徵集 (local detail feature set, LDFS)。接著，運算裝置 120 係依據各類型的本地物件識別碼清單對各類型的感測資料相應的本地細緻特徵集融合為各本地物件之一本地融合特徵 (local fusion feature, LFF)。在一些實施例中，本地物件特徵抽取及融合模組 1312 更執行了各類型之感測資料的上下文獲取處理及上下文融合處理以產生一融合上下文區域。

【0031】本地物件辨識模型 1313 係執行本地物件身分辨識以產生各類型之感測資料相應的本地身分識別清單。舉例來說，運算裝置 120 係將本地物件特徵抽取及融合模組 1312 所產

生之各類型的本地融合特徵輸入本地物件辨識模型 1313 以執行一本地物件身分辨識處理，並將辨識結果以一本地身分識別碼(local identity ID, LIID)標示，再匯集各本地身分識別碼以產生一本地身分識別清單(LIID list)。在一實施例中，本地物件辨識模型 1313 執行本地物件辨識後可產生本地辨識結果及相應的信心度，本地物件辨識模型 1313 可透過一回饋路徑 1314 將所產生的各類型的本地辨識結果及相應的信心度回饋至本地物件偵測及對應模組 1311，使得本地物件偵測及對應模組 1311 可依據相應類型的本地物件辨識結果進行自我學習(self-learning)。

【0032】 因此，本地物件辨識模組 131 所產生的本地物件特徵資訊包括各類型的感測資料之本地物件識別碼清單、本地粗糙特徵集、本地融合特徵、及本地身分識別清單。

【0033】 特徵融合模組 132 係用以依據該本地物件特徵資訊進行一全局物件處理以產生一全局物件特徵資訊。舉例來說，特徵融合模組 132 係包括一全局物件及特徵集對應模組 1321、一上下文區域分析模組 1322、一加權參數選擇模組 1323、及一全局細緻特徵融合模組 1324。全局物件處理例如包括全局物件及特徵集對應模組 1321、上下文區域分析模組 1322、加權參數選擇模組 1323、及全局細緻特徵融合模組 1324 所執行關於全局物件的各種處理。

【0034】 全局物件及特徵集對應模組 1321 係執行一全局物件對應(global object correspondence, GOC)處理及一全局細緻特徵對應(global detail feature correspondence, GDFC)處理以

產生一全局物件識別碼清單及相應的全局細緻特徵集。上下文區域分析模組 1322 係對本地物件特徵抽取及融合模組 1312 所產生之各類型之感測資料的融合上下文區域進行本地上下文分析，並將各類型之感測資料的本地上下文分析結果合併以產生一本地上下文合併結果。

【0035】加權參數選擇模組 1323 係依據上下文區域分析模組 1322 所產生的本地上下文合併結果以判斷要使用鄰近可區分度加權係數或自適應加權係數以進行全局細緻特徵融合處理 (Global Detail Feature Fusion, GDFF)。全局細緻特徵融合模組 1324 係依據來自加權參數選擇模組 1323 所輸出的加權參數以執行全局細緻特徵融合 (global detail feature fusion, GDFF) 處理，例如是將全局物件及特徵集對應模組 1321 所產生的全局細緻特徵集融合為一全局融合特徵。

【0036】因此，特徵融合模組 132 所產生的全局物件特徵資訊包括：全局物件識別碼清單及相應的全局細緻特徵集、及全局融合特徵。

【0037】全局辨識模組 133 係對特徵融合模組 132 所產生的該全局物件特徵資訊進行一全局物件辨識以產生一全局辨識結果。舉例來說，全局辨識模組 133 係將全局細緻特徵融合模組 1324 所產生的全局融合特徵輸入一全局物件辨識模型以辨識各全局融合特徵的全局身分，例如可建立記錄各全局融合特徵的全局身份識別碼的一全局身分識別清單。此外，全局身分識別清單更記錄了全局物件辨識模型的全局辨識結果及其信心度。

【0038】全局辨識模組 133 更可透過一回饋路徑 1331 將其所產生的辨識結果及其信心度回饋至本地物件辨識模型 1313。此外，全局辨識模組 133 更將全局細緻特徵融合模組 1324 所產生的全局融合特徵進行拆解以得到各類型之感測資料的本地細緻特徵，並將所得到的各類型之感測資料的本地細緻特徵回饋至本地物件辨識模型 1313 中之相應類型的本地物件辨識模型，藉以增進各本地物件辨識模型 1313 進行本地物件辨識的準確度，其中回饋路徑 1331 可稱為協同學習 (co-learning)。

【0039】為了便於說明，在後述實施例中主要是以攝影機 110A 搭配麥克風 110B 之配置為例，其他類型之感測器可以用類似的方式運作，並與攝影機 110A 及麥克風 110B 搭配使用。

【0040】第 3A 及 3B 圖係顯示依據本發明一實施例中用於視頻物件之本地物件對應處理、本地細緻特徵融合處理、與本地物件身份辨識的流程圖。

【0041】在步驟 S302，利用複數個攝影機 110A 分別擷取多筆視頻資料。

【0042】在步驟 S304，執行本地物件偵測 (local object detection, LOD) 以判斷各視頻資料中是否有需要被關注的視頻物件，並在步驟 S306，判斷是否發現值得關注的視頻物件。若發現值得關注的視頻物件，則執行步驟 S308 以記錄相應的視頻物件。若沒有發現值得關注的視頻物件，則回到步驟 S302。

【0043】舉例來說，運算裝置 120 例如可在各視頻資料之視訊框內的一空間探索區域中偵測是否有需要被關注的視頻物件。在一些實施例中，運算裝置 120 係由各視頻資料偵測特定

物件，例如人物、人臉、手、汽車、槍、刀、棍等等，但本發明之實施例並不限於此。運算裝置 120 亦可由各視頻資料偵測特定行為，例如聚集、追逐、搶奪、鬥毆、跌倒等行為，但本發明之實施例並不限於此。意即，運算裝置 120 可判斷上述特定物件或特定行為係屬於需關注的視頻物件。

【0044】在一實施例中，運算裝置 120 由各視頻資料偵測不同的特定行為時會有不同的判斷方式。以聚集行為為例，運算裝置 120 係由攝影機 110A 所擷取之視頻資料中判斷在某個空間區域中是否有超過一預定人均密度的情況且持續超過一預定時間，例如可判斷在 5 平方公尺的區域內中的每個平方公尺之範圍內有 3 至 5 人，且持續了 10~30 分鐘且沒有彼此移動之趨向等條件。

【0045】以追逐行為為例，運算裝置 120 係由攝影機 110A 所擷取之視頻資料中判斷出人與人之間的運動軌跡及速率，當兩個人物的運動軌跡類似且速度維持超過一預定速度時，運算裝置 120 可判斷在視頻資料中有追逐行為發生。以跌倒行為為例，運算裝置 120 係由攝影機 110A 所擷取之視頻資料中判斷人物之四肢的位置是否落下的角速度大於一預定角速度並且停留一預定時間。此外，監控系統 100 亦可取得在使用者身上所穿的穿戴式裝置 (wearable device) 所偵測的感測資料，藉以判斷在視頻資料中是否有跌倒行為發生。以搶奪或鬥毆行為為例，運算裝置 120 例如可使用習知的行為判斷演算法以分析視頻資料中是否有搶奪或鬥毆行為發生。

【0046】詳細而言，在人工智慧 (artificial intelligence) 及電

腦視覺 (computer vision) 領域中，行為偵測是個非常高階的應用，除了物件辨識之外，還得運用動態時序、物件移動軌跡、物件交互、物件分佈、密度等關係之資訊。本發明可融合各類型的感測器，以達到互補與全局性的執行物件辨識，且視頻資料僅是本發明中之監控系統 100 中之其中一種感測資料。本發明並不限定於在上述實施例中所揭示的由視頻資料進行不同行為偵測之技術。

【0047】此外，運算裝置 120 可進一步計算所偵測到之各視頻物件的世界座標定位。舉例來說，運算裝置 120 可取得各攝影機 110A 的設置位置 (例如 GPS 座標)、拍攝角度、視角等資訊，並計算在各攝影機 110A 所拍攝之視頻資料中的各視頻物件的世界座標定位資訊。各攝影機 110A 在擷取場景之視訊影像時，亦可加入相應的時間戳記 (time stamp)，以利後續的本地物件對應處理 (local object correspondence, LOC) 及全局物件對應處理 (global object correspondence, GOC)。

【0048】在步驟 S310，判斷是否已偵測出在各視頻資料中所有視頻物件。若是，則執行步驟 S312；若否，則回到步驟 S304。

【0049】在步驟 S312，匯集並分析被偵測出的所有視頻物件。舉例來說，在各視頻資料中可能會偵測出一或多個視頻物件，且各視頻物件亦帶有相應的世界座標定位資訊及時間戳。因此，運算裝置 120 可依據各視頻物件相應的世界座標定位資訊及時間戳來判斷在不同視頻資料中之各視頻物件彼此之間是否相關。

【0050】在步驟 S314，依據被偵測出之各視頻物件相應的世界座標定位資訊(及/或時間戳)以進行一本地物件對應處理(local object correspondence, LOC)，藉以將與同一本地視頻物件相關的多個視頻物件進行對應及聯結，並將同一本地視頻物件相關的多個視頻物件標示相應的一本地物件識別碼(local object ID, LOID)。以本地視頻物件為例，其本地物件識別碼可稱為本地視頻物件識別碼(local video object ID, LVOID)。

【0051】在步驟 S316，建立一第一本地物件識別碼清單(LOID list)，其中第一本地物件識別碼清單記錄了帶有不同本地視頻物件識別碼之一或多本地視頻物件。

【0052】在步驟 S318，依據第一本地物件識別碼清單，匯集並建立各本地視頻物件的一本地粗糙特徵集(local rough feature set, LFRS)，其中本地粗糙特徵集包括各本地視頻物件的方向、距離、概略、與結構等資訊。

【0053】在步驟 S320，依據第一本地物件識別碼清單，依序對各本地視頻物件相關的各視頻資料進行一本地細緻特徵抽取處理以建立各本地視頻物件相關的各視頻資料之一本地細緻特徵集(local detail feature set, LDFS)。

【0054】在步驟 S322，依據第一本地物件識別碼清單，執行一本地細緻特徵融合(local detail feature fusion, LDFE)處理以依序將各本地視頻物件相關的各視頻資料所相應的本地細緻特徵集融合為各本地視頻物件之一本地融合特徵(local fusion feature, LFF)。

【0055】在步驟 S324，將各本地視頻物件之本地融合特徵

輸入至一本地物件辨識(local object recognition, LOR)模型以執行一本地物件身分辨識處理，並將辨識結果以一本地身分識別碼標示，再匯集各本地身分識別碼(local identity ID, LIID)以產生一本地身分識別清單(LIID list)L1。其中在第3A-3B圖之流程中所產生的各本地身分識別碼係標示在視頻資料中相應的各本地視頻物件，故亦可稱為本地視頻身分識別碼(local video identity ID, LVIID)，且本地身分識別清單L1亦可稱為本地視頻身分識別清單。

【0056】 第4A及4B圖係顯示依據本發明一實施例中用於音頻物件之本地物件對應處理、本地細緻特徵融合處理、與本地物件身份辨識的流程圖。

【0057】 在步驟S402，利用複數個麥克風110B分別擷取多筆音頻資料。

【0058】 在步驟S404，執行本地物件偵測(local object detection, LOD)以判斷各音頻資料中是否有需要被關注的音頻物件(audio object)，並在步驟S406，判斷是否發現值得關注的音頻物件。若發現值得關注的音頻物件，則執行步驟S408以記錄相應的音頻物件。若沒有發現值得關注的音頻物件，則回到步驟S402。

【0059】 舉例來說，運算裝置120例如可在各音頻資料之一音頻段內的一時間探索區域中偵測是否有需要被關注的音頻物件。在一些實施例中，運算裝置120係由各音頻資料偵測特定物件聲響或事件聲響，例如槍聲、爆炸聲、哭聲、吵鬧聲、撞擊聲等等，但本發明之實施例並不限於此。意即，運算裝置

120可判斷上述特定物件聲響或事件聲響即為現實環境中的異常聲音，故屬於需關注的音頻物件。對上述異常聲音進行特徵提取例如可使用傳統的語音信號處理方法如梅爾頻率倒譜系數(Mel-Frequency Cepstrum Coefficient、MFCC)、線性預測倒譜系數(Linear Prediction Cepstrum Coefficient、LPCC)等。

【0060】然而，現實環境中還存在許多其他類型的聲音，例如：車輛鳴笛聲、腳步聲，還有其他低頻大氣雜訊等等，且傳統的語音信號處理方法並無法處理上述其他類型的聲音。在一實施例中，運算裝置120係將各麥克風110B所擷取的音頻信號轉換為異常聲音的聲譜圖，並採用2D-Gabor濾波器對聲譜圖時頻特徵進行特徵描述。接著，運算裝置120採用隨機非負獨立成分分析(Stochastic Non-negative Independent Component Analysis，SNICA)提取異常聲音的聲譜圖特徵，再採用稀疏表示分類(Sparse Representation Classification，SRC)方法進行分類識別，藉以判斷出現實環境中之其他類型的異常聲音。

【0061】此外，運算裝置120可進一步計算所偵測到之各音頻物件的世界座標定位。舉例來說，運算裝置120可取得各麥克風110B的設置位置(例如世界座標)、收音角度、間隔距離等資訊，並計算在各麥克風110B所接收之音頻資料中的各音頻物件的世界座標定位資訊。各麥克風110B在擷取場景之音頻資料時，亦可加入相應的時間戳記(time stamp)，以利後續的本地物件對應處理(local object correspondence，LOC)及全局物件對應處理(global object correspondence，GOC)。

【0062】在步驟S410，判斷是否已偵測出在各音頻資料中

所有音頻物件。若是，則執行步驟 S412；若否，則回到步驟 S404。

【0063】 在步驟 S412，匯集並分析被偵測出的所有音頻物件。舉例來說，在各音頻資料中可能會偵測出一或多個音頻物件，且各音頻物件亦帶有相應的世界座標定位資訊及時間戳。因此，運算裝置 120 可依據各音頻物件相應的世界座標定位資訊及時間戳來判斷在不同音頻資料中之各音頻物件彼此之間是否相關。

【0064】 在步驟 S414，依據被偵測出之各音頻物件相應的世界座標定位資訊(及/或時間戳)以進行一本地物件對應處理(local object correspondence, LOC)，藉以將與同一本地音頻物件相關的多個音頻物件進行對應及聯結，並將同一本地音頻物件相關的多個視頻物件標示相應的一本地物件識別碼(local object ID, LOID)。以本地音頻物件為例，其本地物件識別碼可稱為本地音頻物件識別碼(local audio object ID, LAOID)。

【0065】 在步驟 S416，建立一第二本地物件識別碼清單，其中第二本地物件識別碼清單記錄了帶有不同本地音頻物件識別碼之一或多本地音頻物件。

【0066】 在步驟 S418，依據第二本地物件識別碼清單，匯集並建立各本地視頻物件的一本地粗糙特徵集(local rough feature set, LFRS)，其中本地粗糙特徵集包括各本地音頻物件的方向、距離、概略、與結構等資訊。

【0067】 在步驟 S420，依據第二本地物件識別碼清單，依序對各本地音頻物件相關的各音頻資料進行一本地細緻特徵

抽取處理以建立各本地音頻物件相關的各音頻資料之一本地細緻特徵集(local detail feature set, LDFS)。各本地音頻物件的本地細緻特徵集例如包括各本地音頻物件之音量、音高、音色等音頻細緻特徵。

【0068】在步驟 S422，依據第二本地物件識別碼清單，執行一本地細緻特徵融合處理以依序將各本地音頻物件相關的各音頻資料所相應的本地細緻特徵集融合為各本地音頻物件之一本地融合特徵(local fusion feature, LFF)。

【0069】在步驟 S424，將各本地音頻物件之本地融合特徵輸入至一本地物件辨識(local object recognition, LOR)模型以執行一本地物件身分辨識處理，並將辨識結果以一本地身分識別碼(local identity ID, LIID)標示，再匯集各本地身分識別碼以產生一本地身分識別清單(LIID list)L2。其中在第4A-4B圖之流程中所產生的各本地身分識別碼係標示在音頻資料中相應的各音頻物件，故亦可稱為本地音頻身分識別碼，且本地身分識別清單L2亦可稱為本地音頻身分識別清單。

【0070】第5圖係顯示依據本發明一實施例中之全局物件對應處理及建立全局細緻特徵集之流程圖。在一實施例中，在第5圖中全局物件對應處理及建立全局細緻特徵集之流程係使用第3A-3B圖及第4A-4B圖中關於視頻物件及音頻物件之各種處理所產生之資訊及物件清單。

【0071】在步驟 S502，逐一比對第一本地物件識別碼清單中之各本地視頻物件及第二本地物件識別碼清單中的各本地音頻物件的時間戳記。

【0072】在步驟 S504，判斷本地視頻物件及本地音頻物件的時間戳記是否吻合。若時間戳記吻合，則執行步驟 S506；若時間戳記不吻合，則執行步驟 S508。

【0073】在步驟 S506，將本地視頻物件之第一本地粗糙特徵集及第一世界座標定位資訊與本地音頻物件之第二本地粗糙特徵集及第二世界座標定位資訊進行比對。

【0074】在步驟 S508，判斷在第一本地物件識別碼清單中之各本地視頻物件及第二本地物件識別碼清單中的各本地音頻物件是否比對完畢。若是，則執行步驟 S514；若否，則回到步驟 S502。

【0075】在步驟 S510，判斷第一本地粗糙特徵集與第二本地粗糙特徵集以及世界座標定位資訊是否吻合。若是，則執行步驟 S512；若否，則執行步驟 S508。舉例來說，若步驟 S510 的判斷結果為「是」，則表示在第一本地物件識別碼清單中所選擇的本地視頻物件與在第二本地物件識別碼清單中所選擇的本地音頻物件的時間戳記吻合，且兩者相應的本地粗糙特徵集及世界座標定位資訊皆吻合。因此，運算裝置 120 可判斷此本地視頻物件及此本地音頻物件與同一物件有關。

【0076】在步驟 S512，記錄對應成功的本地視頻物件及本地音頻物件，並建立一全局物件識別碼清單(global object identity list)及全局粗糙特徵集(global rough feature set)。舉例來說，對應成功的本地視頻物件及本地音頻物件即可互相聯結且可視為一全局物件(global object)，且運算裝置 120 係指派一全局物件識別碼(global object id, GOID)至該全局物件。因此，

運算裝置 120 可將各全局物件及相應的全局物件識別碼記錄於全局物件識別碼清單中。此外，對應成功的本地視頻物件及本地音頻物件亦分別有相應的本地粗糙特徵集，因此運算裝置 120 亦將此本地視頻物件的本地粗糙特徵集及此本地音頻物件的本地粗糙特徵集互相聯結以形成該全局物件的全局粗糙特徵集(global rough feature set, GRFS)。

【0077】在步驟 S514，將全局物件識別碼清單中之各全局物件所相應之不同類型的本地物件識別碼清單及相應的本地融合特徵合併成全局物件識別碼清單所屬的全局細緻特徵集(global detail feature set, GDFS)。舉例來說，在全局物件識別碼清單中已包括對應成功的本地視頻物件及本地音頻物件，且對應成功的本地視頻物件及本地音頻物件亦分別有相應的本地融合特徵(LFF)，因此運算裝置 120 亦將此本地視頻物件的本地融合特徵及此本地音頻物件的本地融合特徵互相聯結以產生該全局物件相應的全局細緻特徵集。

【0078】第 6A 圖係顯示依據本發明一實施例中利用多個攝影機對一場景擷取視頻資料的示意圖。

【0079】舉例來說，若監控系統 100 在場景 600 中配置 4 個攝影機 110A 用以分別擷取場景 600 中之視頻資料。在場景 600 中有三個物件，例如人物 1(即物件 O1)、人物 2(即物件 O2)、及人物 3(即物件 O3)，分別為一男性大人、一男性兒童、及一女性大人，如第 6A 圖所示。4 個攝影機 110A 例如為分別安裝於不同位置之攝影機 110A-1、110A-2、110A-3、及 110A-4。在此情境中，因為物件被遮蔽或拍攝角度的關係，這三個物件 O1、O2、及

O3可能只被部分的攝影機所拍攝到。例如物件O1僅被攝影機110A-1、110A-2、及110A-3拍攝到，物件O2僅被攝影機110A-1及110A-2拍攝到，且物件O3僅被攝影機110A-1、110A-2、及110A-4拍攝到。運算裝置120會先針對於各個攝影機所拍攝到的內容進行物件偵測，例如可偵測出視頻物件VO1、VO2、VO3。運算裝置120並將對各視頻物件VO1、VO2、VO3指派相應的一物件識別碼(Object Identifier, OID)進行標示，例如視頻物件VO1、VO2、VO3所相應的物件識別碼分別為VOID1、VOID2、及VOID3。其中，視頻物件VO1、VO2、VO3所相應的物件識別碼VOID1、VOID2、及VOID3即可稱為視訊資料的本地物件識別碼(Local Object ID, LOID)。

【0080】對於物件O1來說，因為物件O1僅被攝影機110A-1、110A-2、及110A-3拍攝到，故運算裝置120接著對攝影機110A-1、110A-2、110A-3之視頻資料進行本地細緻特徵抽取處理以取得其本地視頻細緻特徵(例如物件之顏色、紋理、及形狀等等)，例如可分別得到關於物件O1有關的本地視頻細緻特徵集VidF1_O1、VidF2_O1、VidF3_O1。運算裝置120接著對本地視頻細緻特徵集VidF1_O1、VidF2_O1、VidF3_O1執行本地細緻特徵融合處理，可得到有關於物件O1之融合視頻細緻特徵VidFF_O1。簡單來說，融合視頻細緻特徵VidFF_O1可表示在不同角度之攝影機110A所拍攝到之同一物件O1的不同視頻特徵。

【0081】類似地，對於物件O2來說，因為物件O2僅被攝影機110A-1及110A-2拍攝到，故運算裝置120對攝影機110A-1及

110A-2之視頻資料進行本地細緻特徵抽取處理以取得其視頻細緻特徵(例如物件之顏色、紋理、及形狀等等)，例如可分別得到物件 O2 有關的本地視頻細緻特徵集 VidF1_O2、VidF2_O2。運算裝置 120 接著對本地視頻細緻特徵集 VidF1_O2、VidF2_O2 執行本地細緻特徵融合處理後會得到關於物件 O2 的融合視頻細緻特徵 VidFF_O2。融合視頻細緻特徵 VidFF_O2 可表示在不同角度之攝影機 110A 所拍攝到之同一物件 O2 的不同視頻特徵。

【0082】 類似地，對於物件 O3 來說，因為物件 O3 僅被攝影機 110A-1、110A-2、及 110A-4 拍攝到，故運算裝置 120 係對攝影機 110A-1、110A-2、及 110A-4 之視頻資料進行本地細緻特徵抽取處理以取得其視頻細緻特徵(例如物件之顏色、紋理、及形狀等等)，例如可分別得到關於物件 O3 的本地視頻細緻特徵集 VidF1_O3、VidF2_O3、及 VidF4_O3。運算裝置 120 接著對本地視頻細緻特徵集 VidF1_O3、VidF2_O3、及 VidF4_O3 執行本地細緻特徵融合處理後會得到關於物件 O3 的融合視頻細緻特徵 VidFF_O3。融合視頻細緻特徵 VidFF_O3 可表示在不同角度之攝影機 110A 所拍攝到之同一物件 O3 的不同視頻特徵。

【0083】 需注意的是，攝影機 110A-1、110A-2、110A-3、及 110A-4 在擷取場景之視訊影像時，即已加入相應的時間戳記 (time stamp)，且融合視頻細緻特徵 VidFF_O1、VidFF_O2、VidFF_O3 亦帶有相應的時間戳記。舉例來說，第一本地物件識別碼清單包括融合視頻細緻特徵 VidFF_O1、VidFF_O2、VidFF_O3 及相應的時間戳記。

【0084】 運算裝置 120 係將融合視頻細緻特徵 VidFF_O1、VidFF_O2、VidFF_O3 輸入一本地視頻物件辨識模型以辨識融合視頻細緻特徵 VidFF_O1、VidFF_O2、VidFF_O3 之每一者所相應的身分，例如運算裝置 120 可分別指派一本地視頻身分識別碼 (local video identity identifier, LVIID) 至融合視頻細緻特徵 VidFF_O1、VidFF_O2、VidFF_O3，例如分別為本地視頻身分識別碼 LVIID1、LVIID2、及 LVIID3。其中，運算裝置 120 係將本地視頻身分識別碼 LVIID1、LVIID2、及 LVIID3 記錄於一本地視頻身份識別清單 (例如本地身份識別清單 L1) 中。

【0085】 第 6B 圖係顯示依據本發明一實施例中利用多個麥克風對一場景擷取音訊資料的示意圖。

【0086】 若監控系統 100 在場景 600 中更配置了 3 個麥克風 110B 用以分別擷取場景 600 中之音頻資料，如 6B 圖所示。3 個麥克風 110B 例如為分別安裝於不同位置之麥克風 110B-1、110B-2、及 110B-3。其中麥克風 110B-1、110B-2、及 110B-3 例如可分別依附於第 6A 圖中之攝影機 110A-1、110A-2、及 110A-3 以擷取音頻資料，亦可分別設置於在場景 600 中之不同位置以擷取音頻資料。

【0087】 在此情境中，因為物件被遮蔽、音量、或距離的關係，這三個物件 O1、O2、及 O3 所發出的聲音可能只被部分的麥克風 110B 所接收到。例如物件 O1 的聲音僅被麥克風 110B-2、及 110B-3 所擷取到，物件 O2 的聲音僅被麥克風 110B-1、及 110B-3 所擷取到，且物件 O3 的聲音僅被麥克風 110B-3 所擷取到。

【0088】對於物件 O1 來說，因為物件 O1 僅被麥克風 110B-2 及 110B-3 所擷取到，故運算裝置 120 接著對麥克風 110B-2 及 110B-3 所擷取的音頻資料進行本地細緻特徵抽取處理以取得其本地音頻細緻特徵(例如音量、音調、音色等等)，例如可得到關於物件 O1 有關的本地音頻細緻特徵集 AudF2_O1、AudF3_O1。對本地音頻細緻特徵集 AudF2_O1、AudF3_O1 在執行本地細緻特徵融合後會得到有關於物件 O1 之融合音頻細緻特徵 AudFF_O1。簡單來說，融合音頻細緻特徵 AudFF_O1 可表示在不同位置之麥克風 110B 所擷取到之同一物件 O1 的不同音頻特徵。

【0089】類似地，對於物件 O2 來說，因為物件 O2 僅被麥克風 110B-1 及 110B-3 所擷取到，故運算裝置 120 接著對麥克風 110B-1 及 110B-3 所擷取的音頻資料進行本地細緻特徵抽取處理以取得其音頻細緻特徵(例如音量、音調、音色等等)，例如可得到關於物件 O2 有關的本地音頻細緻特徵集 AudF1_O2 及 AudF3_O2。對本地音頻細緻特徵集 AudF1_O2 及 AudF3_O2 執行本地細緻特徵融合後會得到融合音頻細緻特徵 AudFF_O2。融合音頻細緻特徵 AudFF_O2 可表示在不同位置之麥克風 110B 所接收到之同一物件 O2 的不同音訊特徵。

【0090】類似地，對於物件 O3 來說，因為物件 O3 僅被麥克風 110B-3 所擷取到，故運算裝置 120 接著對麥克風 110B-3 所擷取的音頻資料進行本地細緻特徵抽取處理以取得其音頻細緻特徵(例如音量、音調、音色等等)，例如可得到關於物件 O3 有關的本地音頻細緻特徵集 AudF3_O3。對本地音頻細緻特徵集

AudF3_O3 執行本地細緻特徵融合後，會得到融合音頻細緻特徵 AudFF_O3。在此實施例中，融合音頻細緻特徵集 AudFF_O3 即等於音頻細緻特徵集 AudF3_O3。融合音頻細緻特徵 AudFF_O3 可表示在不同位置之麥克風 110B 所接收到之同一物件 O3 的不同音訊特徵。

【0091】需注意的是，麥克風 110B-1、110B-2、110B-3 在擷取場景 600 之音頻信號時，亦加入相應的時間戳記 (timestamp)，且融合音頻細緻特徵 AudFF_O1、AudFF_O2、AudFF_O3 均帶有相應的時間戳記。意即，本地音頻物件清單包括融合音頻細緻特徵 AudFF_O1、AudFF_O2、AudFF_O3 及相應的時間戳記。

【0092】運算裝置 120 係將融合音頻細緻特徵 AudFF_O1、AudFF_O2、AudFF_O3 輸入一本地音頻物件辨識模型以辨識融合音頻細緻特徵 AudFF_O1、AudFF_O2、AudFF_O3 之每一者所相應的身分，例如運算裝置 120 可分別指派一本地音頻身分識別碼 (local audio identity identifier, AIID) 至融合音頻細緻特徵 AudFF_O1、AudFF_O2、AudFF_O3，例如分別為 LAIID1、LAIID2、及 LAIID3。其中，運算裝置 120 係將本地音頻身分識別碼 LAIID1、LAIID2、及 LAIID3 記錄於一本地音頻身份識別清單 (例如本地身份識別清單 L2) 中。

【0093】第 7A 圖係顯示依據本發明一實施例之視頻幀中之不同空間分割方式的示意圖。

【0094】如第 7A 圖所示，在每一張視頻幀 (video frame) 中，可分為不同的空間分割方式，以利於運算裝置 120 進行不同的

影像偵測、影像辨識、及影像分析處理。舉例來說，視頻幀 700 可包括不同的區域，例如是感興趣區域 (region of interest, ROI) 710、探索區域 720、及上下文區域 730。感興趣區域 710 例如為視頻物件 715 在視頻幀 700 中的空間範圍。探索區域 720 係表示視頻物件 715 所屬的感興趣區域 710 的周圍，意即在電腦視覺中用於追蹤 (Tracking) 視頻物件 715 時所運作的範圍。上下文 (context) 區域 730 則會比探索區域 720 大，且上下文區域 730 是用於視頻物件 715 之上下文分析 (context analysis) 所使用的空間範圍。

【0095】第 7B 圖係顯示依據本發明一實施例之音頻段中之不同時間分割方式的示意圖。

【0096】如第 7B 圖所示，在每一個音頻段 (audio segment) 中，可分為不同的時間分割方式，以利於運算裝置 120 進行不同的音頻偵測、音頻辨識、及音頻分析處理。舉例來說，音頻段 750 可包括不同的區域，例如是感興趣區域 (region of interest, ROI) 760、探索區域 770、及上下文區域 780。感興趣區域 760 例如為音頻物件 755 在音頻段 750 中的時間範圍。探索區域 770 係表示音頻物件 755 所屬的感興趣區域 760 的鄰近時間範圍，意即在運算裝置 120 中用於追蹤 (Tracking) 音頻物件 755 時所運作的範圍。上下文 (context) 區域 780 則會比探索區域 770 大，且上下文區域 780 是用於音頻物件 715 之上下文分析 (context analysis) 所使用的時間範圍。

【0097】從第 7A 及 7B 圖之實施例，可得知感興趣區域可為物件所在的空間區域或時間區域。當運算裝置 120 要追蹤物件

時，則採用在空間及時間上比感興趣區域(例如感興趣區域 710 及 760)大探索區域(例如探索區域 720 及 770)。此外，當運算裝置 120 欲進行上下文分析時，運算裝置 120 會採用在空間及時間上比探索區域(例如探索區域 720 及 770)大的上下文區域，例如第 7A 圖之上下文區域 730 或第 7B 圖之上下文區域 780。

【0098】詳細而言，上下文區域(Context)的功能是定義一個探索的邊界做為最大探索區域(Exploration Region)，而最小探索區域即為感興趣區域(ROI)。當運算裝置 120 執行物件追蹤時，會先由用戶或者由運算裝置 120 自行定義一個預測感興趣區域(Predicted ROI)，再由辨識模型(例如本地物件辨識模型 1313)於探索區域內搜索該物件。在一實施例中，運算裝置 120 例如可設定探索區域之範圍為感興趣區域的兩倍(非限定)。在另一實施例中，運算裝置 120 亦可根據物件的運動速度與方向自動調整探索區域的大小。然而，但為了顧及運算與反應效率，運算裝置 120 通常不會設定過大的探索區域，且用戶可以自己設定上下文區域(Context)加以限制。

【0099】在一實施例中，在運算裝置 120 要執行全局細緻特徵融合之前，運算裝置 120 會先執行上下文分析(例如上下文區域分析模組 1322)計算出各特徵的權重分佈。若運算裝置 120 依據上下文分析之結果判斷各類型之本地融合特徵(LFF)的權重分佈係偏向某些本地融合特徵時(例如可視為本地融合特徵的差異很大)，這將導致全局物件辨識之結果產生偏差，此時需採用自適應加權係數(Adaptive Weighting Coefficient, AWC)來進行特徵加權。反之，若運算裝置 120 依據上下文分析之結

果判斷各類型之本地特徵的權重差異不大，此時各類型之本地特徵的權重可稱為鄰近可區分度加權係數 (Adjacent Differential Weighting Coefficient, ADWC)。若運算裝置 120 針對預測感興趣區域執行上下文分析時，運算裝置 120 在計算不同類型的特徵值之差異量時的計算範圍即為預測感興趣區域，此預測感興趣區域亦可稱之為感興趣上下文區域 (Interested Context)。此外，在經過辨識模型對其探索區域進行搜索與辨識之後以確認物件所存在的範圍之感興趣區域，則可稱為已辨識感興趣區域 (Recognized ROI)。

【0100】 第 8A 圖係顯示依據本發明一實施例中之依據上下文分析處理以選擇全局細緻特徵融合之係數的流程圖。

【0101】 在步驟 S802，執行一全局上下文分析處理。舉例來說，全局上下文分析處理係分析各視頻物件及各音頻物件所相應的一上下文區域。

【0102】 在步驟 S804，判斷是否適合採用鄰近可區分度加權係數 (ADWC)。步驟 S804 亦可稱為權重判斷步驟。若判斷適合採用鄰近可區分度加權係數 (ADWC) (例如本地融合特徵差異不大)，則執行步驟 S806，採用鄰近可區分度加權係數 (ADWC) 進行全局細緻特徵融合處理；若判斷不適合採用鄰近可區分度加權係數 (ADWC) (例如本地融合特徵差異很大)，則執行步驟 S808，採用自適應加權係數 (AWC) 進行全局細緻特徵融合處理。

【0103】 舉例來說，在步驟 S806，運算裝置 120 係採用鄰近可區分度加權係數 (ADWC) 進行全局細緻特徵融合處理。舉例

來說，運算裝置 120 可先更新鄰近可區分度加權係數 (ADWC)，再用更新後的鄰近可區分度加權係數 (ADWC) 進行全局細緻特徵融合處理，藉以將在全局物件識別碼清單中之各不同類型的物件的全局細緻特徵集融合並產生一全局融合特徵 (global fusion feature)。

【0104】在步驟 S808，運算裝置 120 係直接採用自適應加權係數 (AWC) 進行全局細緻特徵融合處理。其中當運算裝置 120 依據上下文區域之分析結果判斷不適合採用鄰近可區分度加權係數 (ADWC) (例如局部特徵明顯)，運算裝置 120 會以一回授路徑將自適應加權係數做為輸入加權係數以進行全局細緻特徵融合處理，意即將在全局物件識別碼清單中之各全局物件所包括的不同類型之物件的全局細緻特徵集融合以產生一全局融合特徵 (global fusion feature)。

【0105】在步驟 S810，將各全局物件相應的全局融合特徵輸入一全局物件辨識模型 (global object recognition model) 以進行全局物件之身分辨識，並產生一全局身份識別清單 (global identity ID list)。舉例來說，運算裝置 120 係可分別指派一全局身分識別碼 (global identity identifier, GIID) 至各全局融合細緻特徵，且運算裝置 120 係將全局身分識別碼記錄於全局身分識別清單中。此外，全局身分識別清單更記錄了全局物件辨識模型的全局辨識結果及其信心度。

【0106】在步驟 S812，將本地細緻特徵、及全局辨識結果及其信心度回饋至各本地物件辨識模型。舉例來說，全局融合特徵除了輸入至全局物件辨識模型之外，更會進一步分解為原

本的本地細緻特徵，並回饋至相應的本地物件辨識模型。

【0107】第 8B 圖係顯示依據本發明第 8A 圖之實施例中全局上下文分析處理及權重判斷步驟的流程圖。舉例來說，第 8B 圖係顯示了第 8A 圖中之步驟 S802 及 S804 之細部流程，其中步驟 S802 例如執行了全局上下文分析處理，步驟 S804 則執行了權重判斷步驟。

【0108】在步驟 S8021，定義一預測感興趣區域(Predicted ROI)。舉例來說，預測感興趣區域亦可稱為感興趣上下文區域(Interested Context)，且可由使用者或是由運算裝置 120 定義。

【0109】在步驟 S8022，執行各類型之本地上下文分析(local context analysis, LCA)處理，並計算各類型之本地細緻特徵之特徵值的差異值，並將上述差異值進行正規化(normalize)。舉例來說，本地上下文分析(LCA)是對當前不同類型(例如視頻、音頻、氣味、味道、觸覺等)之感測資料進行特徵抽取與融合以取得各類型之感測資料的本地融合特徵(LFF)。

【0110】運算裝置 120 係對當前之各類型之感測資料的本地融合特徵(LFF)執行一套特定的計算與分析，以得到當前之各類型之感測資料的本地融合特徵(LFF)相應的權重值。上述權重值的計算方式，例如可針對顏色、紋理、形狀等視頻細緻特徵、以及針對音量、音高、音色等音頻細緻特徵為例進行計算。其餘類型之感測資料的本地融合特徵亦可採用類似的方式計算以取得相應的權重值。

【0111】以視頻細緻特徵為例，顏色特徵可包括疏密、飽

和度、亮度等特徵值的差異值，紋理特徵可包括圖樣(pattern)特徵值的差異值，形狀特徵則可包括線條、相對位置、相對長度、相對方向等特徵值的差異值。以音頻細緻特徵為例，音量特徵可包括聲音能量的差異值，音高特徵可包括聲音頻率的差異量，音色特徵則可包括發聲體的諧音或泛音成分比例的差異值。因為每一個本地細緻特徵的特性彼此不同，還需要將各本地細緻特徵之差異值經過正規化(Normalization)才能對各本地細緻特徵的正規化差異值進行評比。上述之正規化差異值係表示各被選定的本地融合特徵(LFF)在整體評價中的相對重要程度，例如可以用自然數表示，故可能是負數或者是零。

【0112】 在一些實施例中，運算裝置120例如可利用局部二值模式(local binary patterns, LBP)以計算上述不同類型之各本地細緻特徵的差異值。

【0113】 在步驟S8023，合併各類型之所有本地細緻特徵的本地上下文分析結果。舉例來說，在取得各類型之各本地細緻特徵的正規化差異值後，運算裝置120再合併各類型之所有本地細緻特徵的本地上下文分析結果。

【0114】 在步驟S8024，依據各本地細緻特徵之正規化差異值指派各本地細緻特徵相應的一權重值。若有一本地細緻特徵所相應的正規化差異值的數值較大，該本地細緻特徵之權重值也會較大。若有一本地細緻特徵所相應的正規化差異值的數值較小，則該本地細緻特徵之權重值也會較小。

【0115】 在步驟S8025，取得一預定區間的上閾值及下閾值。舉例來說，運算裝置120可取得其類型或應用的實際情況

所定義的權重值之預定區間的上/下閾值 (upper/lower threshold)(例如可由使用者定義或由運算裝置 120 自行定義)。

【0116】在步驟 S8041，判斷各本地細緻特徵相應的權重值是否均在該預定區間。當各本地細緻特徵相應的權重值均在該預定區間時，表示各本地細緻特徵相應的權重值的差異不大，故可執行步驟 S8042，判斷適合採用鄰近可區分度加權係數 (ADWC)。當有任一本地細緻特徵相應的權重值不在該預定區間時(意即有任一本地細緻特徵之正規化差異值超出上閾值或下閾值)，即可認定過於偏向此本地細緻特徵，故可執行步驟 S8043，判斷不適合採用鄰近可區分度加權係數 (ADWC)，意即適合採用自適應加權係數 (AWC)。

【0117】在一實施例中，假設選定三個視訊有效的本地融合特徵 (LFF) 分別稱之為特徵 A、B、C，還有兩個音訊有效的本地融合特徵 (LFF) 分別稱之為特徵 D、E，分別以各本地上下文分析並正規化合併其結果，以獲得特徵 A~E 相應的權重值 $W_A \sim W_E$ ，例如 $W_A=5$ 、 $W_B=2$ 、 $W_C=4$ 、 $W_D=6$ 、 $W_E=3$ 。若預定區間為 3~6，表示下閾值為 3 且上閾值為 6，在此狀況下並沒有偏向某個特徵，故可將權重值 $W_A \sim W_E$ 設定為鄰近可區分度加權係數 (ADWC)，並套用到各特徵以改變各特徵的相對重要程度。但若得到特徵 A~E 相應的權重值 $W_A \sim W_E$ 分別為 $W_A=7$ 、 $W_B=2$ 、 $W_C=4$ 、 $W_D=6$ 、 $W_E=3$ ，當使用下閾值為 3 且上閾值為 6 的預定區間，因為 $W_A=7$ 已經超出上閾值，此即表示已經判斷結果會偏向特徵 A，此時運算裝置 120 則會採用自適應加權係數 (AWC)。

【0118】詳細而言，為了增強物件辨識能力、準確性以及預測能力，運算裝置 120 可依用戶需求採納不同類型之有效特徵資訊，再以全局上下文分析來決定要採用哪一個加權係數，並融合運用所選定的特徵。此外，運算裝置 120 可將全局物件辨識結果回饋給本地物件辨識模型。舉例來說，當運算裝置 120 欲將全局細緻特徵集 (GDFS) 融合成全局融合特徵 (GFF) 前，運算裝置 120 需先挑選出有效的本地融合特徵。所謂的有效或無效特徵指的是對於辨識精確度能否起作用的特徵，如皮膚的紋理對於預測年紀就是有效的特徵，皮膚的顏色對於預測年紀是無效的特徵。意即，針對不同類型的感測資料，使用者可設定那些本地融合特徵在進行全局物件辨識或本地物件辨識時是屬於有效特徵。

【0119】第 8C-1 及 8C-2 圖係顯示依據本發明一實施例中之全局細緻特徵融合及全局物件辨識之流程圖。

【0120】在步驟 S8201，分別設定鄰近可區分度加權係數 (ADWC) 及自適應加權係數 (AWC) 的初始權重值。

【0121】在步驟 S8202，取得追蹤條件，再依追蹤條件定義預測感興趣區域。舉例來說，追蹤條件可由使用者定義，例如可定義為穿特定顏色或花紋的人物、或是移動最快的人物等條件。接著，運算裝置 120 會由不同類型的感測資料中分別定義出預測感興趣區域。

【0122】在步驟 S8203，融合所有本地的上下文及預測感興趣區域。舉例來說，各類型的感測資料的上下文及預測感興趣區域會被合併。若是第一次執行步驟 S8203，則合併後的所產

生的合併感興趣區域可稱為初始感興趣區域(initial ROI)。

【0123】在步驟 S8204，進行全局上下文分析處理。其中，全局上下文分析處理之細節可參考第 8A－8B 圖之流程。

【0124】在步驟 S8205，判斷是否適合採用鄰近可區分度加權係數(ADWC)。步驟 S8205 之細節可參考第 8B 圖中之流程，故不再贅述。當判斷適合採用鄰近可區分度加權係數(ADWC)，則執行步驟 S8206；當判斷不適合採用鄰近可區分度加權係數(ADWC)，則執行步驟 S8210。

【0125】在步驟 S8206，將各類型的感測資料之全局上下文分析所得到的權重值設定為鄰近可區分度加權係數(ADWC)。舉例來說，當各類型之感測資料的各本地融合特徵相應的權重值均在預定區間內時，表示全局物件辨識的結果不會偏向某一特徵，故可將全局上下文分析所得到的權重值設定為鄰近可區分度加權係數(ADWC)。

【0126】在步驟 S8207，套用鄰近可區分度加權係數(ADWC)以執行全局細緻特徵融合(GDFF)處理以建立一全局融和特徵。舉例來說，因為已判斷適合採用鄰近可區分度加權係數(ADWC)，且已將上下文分析所得到的權重值設定為鄰近可區分度加權係數(ADWC)，故運算裝置 120 可依據各類型之各本地融合特徵相應的權重值進行運算，以執行全局細緻特徵融合處理以產生全局融和特徵。

【0127】在步驟 S8208，將全局融和特徵輸入一全局物件辨識模型進行全局物件辨識。其中，全局物件辨識模型例如可為第 9A-2 圖所示的全局物件辨識模型 920。

【0128】在步驟 S8209，依據全局物件辨識之辨識結果產生已辨識感興趣區域之邊界(boundary)。需注意的是，當執行完步驟 S8209，即表示完成當前時間幀(time frame)的全局物件辨識流程，且可進行下一個時間幀的全局物件辨識流程。

【0129】在步驟 S8210，套用自適應加權係數(AWC)以執行全局細緻特徵融合(GDFF)以建立一全局融和特徵。舉例來說，因為已判斷適合採用自適應加權係數(AWC)，且全局上下文分析所得到的權重值可能會偏重某些特徵，故運算裝置 120 套用自適應加權係數(AWC)以執行全局細緻特徵融合(GDFF)以建立一全局融和特徵。

【0130】在步驟 S8211，將全局融和特徵輸入一全局物件辨識模型進行全局物件辨識。其中，全局物件辨識模型例如可為第 9A-2 圖所示的全局物件辨識模型 920。

【0131】在步驟 S8212，依據全局物件辨識之辨識結果產生已辨識感興趣區域之邊界(boundary)。需注意的是，當執行完步驟 S8212，還需執行步驟 S8213 以判斷是否為第一次執行全局上下文分析，若步驟 S8213 之判斷結果為「是」，則執行步驟 S8214；若步驟 S8213 之判斷結果為「否」，則執行步驟 S8215。

【0132】在步驟 S8214，進行前感興趣區域及後感興趣區域之特徵近似度評比。其中，在步驟 S8214 中，因為是第一次執行全局上下文分析，故前感興趣區域係指對當前時間幀執行全局物件辨識前的感興趣區域，後感興趣區域係指對當前時間幀執行全局物件辨識後的感興趣區域。

【0133】在步驟 S8215，進行前感興趣區域及後感興趣區域

之特徵近似度評比。其中，在步驟 S8215 中，因為並非第一次執行全局上下文分析，故前感興趣區域係指對上一個時間幀 (previous time frame) 執行全局物件辨識後的感興趣區域，後感興趣區域係指對當前時間幀執行全局物件辨識前的感興趣區域。

【0134】在步驟 S8214 及 S8215 中所執行的前感興趣區域及後感興趣區域之特徵近似度評比通常採用巴式距離 (Bhattacharyya distance) 計算。因為每一個特徵之特性不同，其特徵值需經過正規化 (Normalization) 才能彼此評比。若在前感興趣區域及後感興趣區域中的同一特徵值的相似度越高，相應的該特徵之權重值也會隨之變大。將上述特徵之權重值進行正規化後即可得到自適應加權係數 (AWC)。

【0135】在步驟 S8216，更新自適應加權係數 (AWC)。需注意的是，更新後的自適應加權係數 (AWC) 是用於處理下一個時間幀的資料。

【0136】在步驟 S8217，將當前時間幀所得到的已辨識感興趣區域之邊界套用至下一個時間幀，以做為下一個時間幀的預測感興趣區域，接著再重複執行步驟 S8203 ~ S8217 之流程。

【0137】第 8D 圖係顯示依據本發明一實施例中之全局細緻特徵融合與全局物件辨識之資料管線的示意圖。請同時參考第 8C-1 ~ 8C-2 圖及第 8D 圖。

【0138】在第 8D 圖中，左邊標示階段 1 至階段 9 係表示全局細緻特徵融合與全局物件辨識之資料管線之不同階段。TF1 至 TF7 係表示時間幀。ROI1 至 ROI7 係分別表示在時間幀 TF1 ~ TF7

的預測感興趣區域，且 Context1 ~ Context7 係表示在時間幀 TF1 ~ TF7 的上下文區域。

【0139】 在階段 1，進行本地上下文融合及預測 ROI 融合。在時間幀 TF1，因為是第一張時間幀，故可直接使用融合上下文區域及融合感興趣區域進行階段 2 的全局上下文分析 (GCA)。鄰近可區分度加權係數 (ADWC) 或自適應加權係數 (AWC) 在第一張時間幀時會設定為預設值 (default)。

【0140】 在階段 3，依據全局上下文分析 (GCA) 之結果判斷要使用何種加權係數 (weighting coefficient, WC)。若判斷使用自適應加權係數 (AWC)，則在第 8D 圖上的階段 3 係標示 A；若判斷使用鄰近可區分度加權係數 (ADWC)，則在第 8D 圖上的階段 3 係標示 AD。

【0141】 在階段 4，依據所選擇的加權係數進行全局細緻特徵融合 (GDFF)，並產生階段 5 的全局融合特徵 (GFF)。

【0142】 在階段 6，依據階段 5 所產生的全局融合特徵以進行全局物件辨識 (GOR)。並在階段 7 中，依據全局物件辨識之辨識結果產生已辨識感興趣區域之邊界。

【0143】 在階段 8，進行前感興趣區域及後感興趣區域之特徵近似度評比。若是第一次執行全局上下文分析 (即時間幀 TF1)，故前感興趣區域係指對當前時間幀執行全局物件辨識前的感興趣區域，後感興趣區域係指對當前時間幀執行全局物件辨識後的感興趣區域。若非第一次執行全局上下文分析，故前感興趣區域係指對上一個時間幀執行全局物件辨識後的感興趣區域，後感興趣區域係指對當前時間幀執行全局物件辨識前

的感興趣區域。

【0144】在階段 9，更新自適應加權係數(AWC)。舉例來說，若在階段 3 之標記為 A，則表示使用自適應加權係數(AWC)，故需在階段 9 更新自適應加權係數(AWC)。若在階段 3 之標記為 AD，則表示使用鄰近可區分度加權係數(ADWC)，故可省略階段 8 及階段 9。例如在時間幀 TF1、TF2、TF4、及 TF5 均判斷使用自適應加權係數(AWC)，故均會執行階段 8 的進行前感興趣區域及後感興趣區域之特徵近似度評比。

【0145】此外，在階段 9 中之當前時間幀所更新的自適應加權係數(AWC)亦會用於下一個時間幀的階段 3。例如在時間幀 TF1 之階段 9 所產生的自適應加權係數(AWC)已更新為 AWC1，故在時間幀 TF2 之階段 3 中的自適應加權係數(AWC)即為 AWC1，依此類推。

【0146】若在階段 3 中決定使用鄰近可區分度加權係數(ADWC)，則在會當前時間幀中更新鄰近可區分度加權係數(ADWC)的數值。例如在時間幀 TF3 之階段 3 中已決定使用鄰近可區分度加權係數(ADWC)，故會將鄰近可區分度加權係數(ADWC)更新為當前時間幀的鄰近可區分度加權係數 ADWC3，依此類推。

【0147】需注意的是，在階段 7 中所得到的已辨識感興趣區域，例如 ROI1a~ROI7a 之邊界會套用至下一個時間幀，例如在時間幀 TF1 所得到的已辨識感興趣區域 ROI1a 之邊界會套用至時間幀 TF2 中的預測感興趣區域 ROI2 之邊界，依此類推。

【0148】第 8E 圖係顯示依據本發明一實施例中之辨識結果

回饋與強化之全局回饋的流程圖。

【0149】在步驟 S832，將全局融合特徵(GFF)傳送至全局物件辨識模型以進行全局物件辨識，並產生一全局辨識結果及相應的信心度。舉例來說，全局物件辨識模型所輸出的全局辨識結果會表示由目前不同類型的感測資料所偵測出的人物(即全局物件)，且信心度愈高，表示全局辨識結果的可信度也愈高。

【0150】在步驟 S834，定義一信心度閾值。舉例來說，使用者可自行設定所需的信心度閾值，或是可由運算裝置 120 自行決定信心度閾值，其中信心度閾值例如可表示對於全局辨識結果所需的最低限度的可信度。

【0151】在步驟 S836，判斷信心度是否低於信心度閾值。若是，則結束此流程；若否，則執行步驟 S838。舉例來說，若全局辨識結果的信心度低於信心度閾值，則表示目前的全局辨識結果的可信度不高，可能需要待感測器之資料更新或是場景中之物件移動後再更新全局辨識結果。若全局辨識結果的信心度不低於信心度閾值，則表示目前的全局辨識結果有一定的可信度。

【0152】在步驟 S838，將全局細緻特徵(GDFF)分解為各本地融合特徵。舉例來說，因為目前的全局辨識結果有一定的可信度，故可將在全局物件辨識模型中用於進行全局物件辨識的全局細緻特徵分解為原本各類型的本地融合特徵。

【0153】在步驟 S840，將全局辨識結果及相應的信心度、及各本地融合特徵回饋至各本地物件辨識模型(例如第 2 圖所示之本地物件辨識模型 1313)。

【0154】經由第 8E 圖之流程，各本地物件辨識模型可利用此回饋路徑以進行本地物件辨識(local object recognition, LOR)模型及全局物件辨識(global object recognition, GOR)模型的協同學習(co-learning)，故可全局性地自動強化本地物件辨識的能力及準確度。

【0155】第 8F 圖係顯示依據本發明一實施例中之辨識結果回饋與強化之本地回饋的流程圖。在第 8E 圖之流程中，主要是針對全局回饋並可進行本地物件辨識模型及全局物件辨識模型的協同學習。此外，本地端亦可進行類似的回饋，此即稱為本地回饋。

【0156】在步驟 S850，將本地融合特徵(LFF)傳送至本地物件辨識模型以進行本地物件辨識，並產生一本地辨識結果及相應的信心度(例如可包含於本地身份識別清單中)。舉例來說，本地物件辨識模型所輸出的本地辨識結果會表示由目前相同類型的感測資料所偵測出的人物(即本地物件)，且信心度愈高，表示本地辨識結果的可信度也愈高。

【0157】在步驟 S852，定義一信心度閾值。舉例來說，使用者可自行設定所需的信心度閾值，或是可由運算裝置 120 自行決定信心度閾值，其中信心度閾值例如可表示對於本地辨識結果所需的最低限度的可信度。用於本地物件辨識之信心度閾值可與用於全局物件辨識之信心度閾值相同或不同。

【0158】在步驟 S854，判斷信心度是否低於信心度閾值。若是，則結束此流程；若否，則執行步驟 S856。舉例來說，若本地辨識結果的信心度低於信心度閾值，則表示目前的本地辨

識結果的可信度不高，可能需要待感測器之資料更新或是場景中之物件移動後再更新本地辨識結果。若本地辨識結果的信心度不低於信心度閾值，則表示目前的本地辨識結果有一定的可信度。

【0159】在步驟 S856，將本地辨識結果及相應的信心度、及各本地融合特徵回饋至各本地物件偵測模型(例如第2圖所示之本地物件偵測及對應模組 1311)。

【0160】經由第 8F 圖之流程，各本地物件偵測模型可利用此回饋路徑以進行本地物件偵測(local object detection, LOD)模型及本地物件辨識(LOR)模型的自我學習(self-learning)，故可自動強化本地物件偵測及辨識的能力及準確度。在一些實施例中，本地物件偵測模型及本地物件辨識模型可選擇性地參考回饋資訊，亦可依據實際情況及需求以自行決定要如何運用回饋資訊。

【0161】第 9A-1 及 9A-2 圖係顯示依據本發明一實施例中之監控方法的方塊圖。

【0162】在一實施例中，在方塊 902-1~902-N，運算裝置 120 係執行一本本地物件偵測與對應處理，其包括一本本地物件偵測處理及一本本地物件對應處理。在方塊 902-1~902-N 之每一者均會分別接收不同類型之感測器的感測資料，例如方塊 902-1 係接收來自一或多個攝影機 110A(例如攝影機 110A-1~110A-4)所拍攝的視頻資料，方塊 902-2 係接收來自一或多個麥克風 110B(例如麥克風 110B-1~110B-3)所接收的音頻資料，方塊 902-N 則接收來自一或多個氣味感測器 110D(例如氣味感測器

110D-1~110D-3)所偵測到的氣味資料。各個方塊 902 會產生各類型之感測資料的本地物件識別碼清單及相應的本地粗糙特徵集。舉例來說，不同類型之感測資料所相應的本地粗糙特徵集例如包括各不同類型之感測資料之本地物件的方向、距離、概略、與結構等資訊。此外，視頻資料及音頻資料建立關於視頻物件及音頻物件的本地物件識別碼清單及相應的本地粗糙特徵集之流程可參考第 3A-3B 圖及第 4A-4B 圖之實施例。

【0163】 在方塊 904-1~904-N，運算裝置 120 係分別執行不同類型之感測資料之一本地細緻特徵抽取及融合處理，其包括一本地細緻特徵抽取處理及本地細緻特徵融合處理。舉例來說，運算裝置 120 係依據本地物件識別碼清單，依序對各本地物件相關之感測資料進行一本地細緻特徵抽取處理以建立各本地物件相關的各感測資料之一本地細緻特徵集。運算裝置 120 更依據本地物件識別碼清單，執行一本地細緻特徵融合處理以依序將各本地物件相關的各感測資料所相應的本地細緻特徵集融合為各本地物件之一本地融合特徵。

【0164】 在方塊 906-1~906-N，運算裝置 120 係分別將將各本地物件之本地融合特徵輸入至一本地物件辨識 (local object recognition, LOR) 模型以執行一本地物件身分辨識處理，並將辨識結果以一本地身分識別碼標示，再匯集各本地身分識別碼 (local identity ID, LIID) 以產生一本地身分識別清單。

【0165】 在方塊 908，運算裝置 120 係執行一全局物件對應處理以產生一全局物件識別碼清單及各全局物件相應的一全局粗糙特徵集。舉例來說，全局物件對應處理之流程例如可參

考第5圖之實施例，惟在此處不只有視頻物件及音頻物件之對應處理，但其流程是類似的。舉例來說，運算裝置120同樣逐一比對在不同類型之感測資料之本地物件識別碼清單中的本地物件之時間戳記。當時間戳記吻合時，運算裝置120再比對時間戳記吻合的各本地物件之世界座標定位資訊。當世界座標定位資訊亦吻合時，運算裝置120再進一步判斷不同類型之感測資料之本地粗糙特徵集是否吻合。當本地粗糙特徵集亦吻合時，運算裝置120則可將對應成功之不同類型之感測資料的本地物件及相應的本地粗糙特徵集互相聯結並建立相應的一全局物件及相應的全局粗糙特徵集，其中各全局物件均具有相應的全局物件識別碼。

【0166】 在方塊910，運算裝置120係執行一全局細緻特徵集對應處理。舉例來說，運算裝置120係將全局物件識別碼清單中之各全局物件所相應之不同類型的本地物件識別碼清單及相應的本地融合特徵合併成全局物件識別碼清單所屬的全局細緻特徵集。因為在全局物件識別碼清單中已包括對應成功之不同類型的本地物件，且對應成功之不同類型的本地物件亦分別有相應的本地融合特徵，故運算裝置120亦將不同類型的本地物件所相應的本地融合特徵互相聯結以產生該全局物件相應的全局細緻特徵集。

【0167】 在方塊912，運算裝置120係執行一全局上下文分析處理。舉例來說，運算裝置120係分析在全局物件識別碼清單中之各全局物件中的不同類型(跨類型)的本地物件所相應的一上下文區域。以視頻物件為例，可參考第7A圖，運算裝置120

係對視頻幀 700 中之上下文區域 730 的空間範圍進行上下文分析。若以音頻物件為例，可參考第 7B 圖，運算裝置 120 則對音頻段 750 中之上下文區域 780 的時間範圍進行上下文分析。

【0168】 運算裝置 120 更判斷在各本地物件中之上下文區域中是否有任一細緻特徵為明顯。例如視頻細緻特徵包括顏色、紋理、形狀等等，且音頻細緻特徵包括音量、音色、音高等等，且上下文分析處理之結果會選擇要採用鄰近可區分度加權係數或自適應加權係數進行後續的全局細緻特徵融合。此外，每個加權係數均對應至一個細緻特徵。因此，以視頻資料及音頻資料為例，總共有 6 個加權係數。

【0169】 在選擇器 914 中，若在方塊 912 是判斷要採用鄰近可區分度加權係數以進行全局細緻特徵融合，則由方塊 910 所產生的本地身分清單、全局物件識別碼清單及相應的全局細緻特徵集會輸入方塊 916，且運算裝置 120 係更新鄰近可區分度加權係數 (ADWC)。

【0170】 在選擇器 914 中，若在方塊 912 是判斷要採用自適應加權係數 (AWC) 以進行全局細緻特徵融合，則由方塊 910 所產生的本地身分清單、全局物件識別碼清單及相應的全局細緻特徵集會直接輸入至方塊 918 以進行全局細緻特徵融合。此外，在方塊 924 中所產生之更新後的自適應加權係數亦會輸入至方塊 918 以進行全局細緻特徵融合。其中，方塊 924 所產生自適應加權係數例如是由上一次的全局物件辨識模型的辨識結果所決定並進行更新。

【0171】 在方塊 918，運算裝置 120 係執行全局細緻特徵融

合處理。如上所述，全局細緻特徵融合處理之輸入參數可為鄰近可區分度加權係數(ADWC)或自適應加權係數(AWC)，視在方塊912之全局上下文分析的結果而定。詳細而言，運算裝置120係將各全局物件相應的全局細緻特徵集再進行一次特徵融合以得到各全局物件相應的一全局融合特徵(global fusion feature, GFF)。

【0172】在方塊920，運算裝置120係將全局融合特徵輸入一全局物件辨識模型以進行全局物件之身分辨識，並產生一全局身分識別清單(global identity ID list)。舉例來說，運算裝置120係可分別指派一全局身分識別碼(global identity identifier, GIID)至各全局融合特徵，且運算裝置120係將全局身分識別碼記錄於全局身分識別清單中。此外，全局身分識別清單更記錄了全局物件辨識模型的辨識結果及其信心度。

【0173】在選擇器922，若在方塊912的判斷結果為使用鄰近可區分度加權係數進行全局細緻特徵融合，則方塊920所輸出的辨識結果及相應的信心度(兩者可稱為回饋資訊)會直接輸出至不同類型之感測資料的本地物件辨識模型，例如方塊906-1~906-N。在一些實施例中，若一特定類型之感測資料的辨識結果相應的信心度小於一預定比例(例如80%)時，選擇器922則不會將特定類型的感測資料之辨識結果及相應的信心度回饋至方塊906-1~906-N中的本地物件辨識模型。在另一些實施例中，方塊906-1~906-N中的本地物件辨識模型亦可自行決定是否要採用回饋資訊。

【0174】在方塊924，運算裝置120係更新自適應加權係

數，例如可依據上一次的全局物件辨識模型的辨識結果進行更新。

【0175】在方塊 926，運算裝置 120 係將全局細緻特徵進行分解，以得到不同類型的本地細緻特徵。需注意的是，在方塊 926 中分解所得到的不同類型的本地細緻特徵會分別輸入至方塊 906-1 ~ 906-N 的本地物件辨識模型。

【0176】因此，方塊 901-6 ~ 906-N 中的各本地物件辨識模型則可依據相應類型的辨識結果及其信心度(來自方塊 920 並經過選擇器 922)、以及相應類型的細緻特徵(來自方塊 926)以調整或更新目前的本地物件辨識模型，使得下一次的物件辨識可以得到更準確的結果。

【0177】第 9B-1 及 9B-2 圖係顯示依據第 9A-1 及 9A-2 圖之實施例之全局上下文分析處理之詳細方塊圖。在一實施例中，在第 9A-1 及 9A-2 圖之方塊 912 中所執行的全局上下文分析處理可參考第 9B-1 及 9B-2 圖之內容。在第 9A-1 及 9A-2 圖中的方塊 904-1 ~ 904-N 的每一者除了執行相應類型之感測資料的本地細緻特徵抽取處理及本地細緻特徵融合處理之外，還會執行相應類型之感測資料的上下文獲取處理及上下文融合處理。運算裝置 120 在執行本地細緻特徵抽取處理的同時，亦會執行上下文獲取處理，如第 9B-1 及 9B-2 圖之方塊 904-1 ~ 904-N 所示。

【0178】詳細而言，以方塊 904-1 為例，由攝影機 110A-1 ~ 110A-3 所擷取的視頻資料 Video1、Video2、Video3 在經過第 9A-1 圖中的方塊 902-1 後，視頻資料 Video1、Video2、Video3 仍會輸入至方塊 904-1，意即不同的攝影機 110A 所拍攝的視頻

資料均會輸入至方塊 904-1 以分別進行本地細緻特徵抽取處理及上下文獲取處理，如第 9B-1 圖中之方塊 904-1A、904-1B、及 904-1C。在方塊 904-1A~904-1C 中所得到的本地細緻特徵集會輸入至 904-1D 以進行本地細緻特徵融合處理以產生關於視頻資料的本地物件識別碼清單及本地融合特徵，並本地物件識別碼清單及本地融合特徵會輸入至方塊 910 進行全局細緻特徵集建立處理。

【0179】 此外，在方塊 904-1A、904-1B、及 904-1C 中所分別執行的上下文獲取處理，例如可參考已辨識出之視頻物件，並進一步在相應的視頻幀中取得上下文區域及預測感興趣區域。在方塊 904-1E，運算裝置 120 係執行一本地上下文融合處理及 ROI 融合處理以將來自方塊 904-1A、904-1B、及 904-1C 的上下文區域及預測感興趣區域分別進行融合，例如可得到一融合上下文區域及一融合感興趣區域。在第 9B-1 及 9B-2 圖中之方塊 904-2 (例如針對音頻資料 Audio1、Audio2、及 Audio3) 至 904-N (例如針對氣味資料 Smell1、Smell2、及 Smell3) 均可對相應類型的感測資料進行處理，並可得到相應類型的融合上下文區域及融合感興趣區域。

【0180】 在方塊 904-1~904-N 所分別得到的融合上下文區域及融合感興趣區域會輸入至方塊 912 中相應的方塊 912-1~912-N 進行一上下文分析處理，並將方塊 912-1~912-N 的本地上下文分析結果傳送至方塊 9120 以進行一上下文分析結果合併處理及全局感興趣區域 (ROI) 合併處理。運算裝置 120 係依據方塊 9120 所產生的上下文合併結果以判斷要使用鄰近可區分

度加權係數或自適應加權係數以進行全局細緻特徵融合處理。

【0181】第 9C 圖係顯示依據本發明一實施例中之利用多維度感測器資料之監控方法的流程圖。在第 9A-1 ~ 9A-2 圖及第 9B-1 ~ 9B-2 圖之流程圖及方塊圖可合併簡化為第 9C 圖之流程。請同時參考第 9C 圖及第 2 圖。

【0182】在方塊 952，利用同類型感測器組取得感測資料。舉例來說，相同類型之感測器所取得的感測資料是傳送至相應的本地物件偵測與對應模組 1311。

【0183】在方塊 954，進行本地物件偵測與對應 (LOD 及 LOC)。舉例來說，本地物件偵測及對應模組 1311 係分別接收來自攝影機 110A、麥克風 110B、味覺感測器 110C、氣味感測器 110D、及觸覺感測器 110E 之感測資料，並進行相應的感測類型之一本地物件偵測及對應處理 (即包括本地物件偵測處理 (LOD) 及本地物件對應處理 (LOC)) 以產生一本地物件識別碼清單 (LOID list) 及本地粗糙特徵集 (LRFS)。

【0184】在方塊 956，進行本地細緻特徵抽取與融合。舉例來說，本地物件特徵抽取及融合模組 1312 係執行一本地特徵抽取及融合處理，其包括一本地細緻特徵抽取處理 (LDFE) 及一本地物件融合處理 (LDFP)。舉例來說，運算裝置 120 係依據本地物件辨識模組 131 所產生的本地物件識別碼清單及本地粗糙特徵集，對各不同類型之感測資料抽取其本地細緻特徵並建立各類型的感測資料相應的本地細緻特徵集。接著，運算裝置 120 係依據各類型的本地物件識別碼清單對各類型的感測資料相應的本地細緻特徵集融合為各本地物件之一本地融合特徵。在

一些實施例中，本地物件特徵抽取及融合模組 1312 更執行了各類型之感測資料的上下文獲取處理及上下文融合處理以產生一融合上下文區域。此外，本地物件特徵抽取及融合模組 1312 還可將各類型之感測資料之感興趣區域融合以產生融合感興趣區域。

【0185】 在方塊 958，進行本地物件辨識(LOR)。舉例來說，本地物件辨識模型 1313 係執行本地物件身分辨識以產生各類型之感測資料相應的本地身分識別清單。運算裝置 120 係將來自方塊 956 之各類型的本地融合特徵輸入本地物件辨識模型 1313 以執行一本地物件身分辨識處理，並將辨識結果以一本地身分識別碼標示，再匯集各本地身分識別碼以產生一本地身分識別清單(LIID list)。在一實施例中，本地物件辨識模型 1313 可透過一回饋路徑(例如箭頭 959)將所產生的各類型的本地物件辨識結果回饋至本地物件偵測及對應模組 1311，使得本地物件偵測及對應模組 1311 可依據相應類型的本地物件辨識結果進行自我學習(self-learning)。

【0186】 在方塊 960，進行全局物件對應(GOC)。舉例來說，全局物件及特徵集對應模組 1321 可依據來自方塊 954 的本地物件識別碼清單(LOID list)及本地粗糙特徵集(LRFS)、及來自方塊 958 之本地身分識別清單(LIID list)進行全局物件對應(GOC)以產生一全局物件識別碼清單(GOID list)及全局粗糙特徵集(GRFS)。

【0187】 在方塊 962，進行全局細緻特徵對應(GDFC)。舉例來說，全局物件及特徵集對應模組 1321 可依據來自方塊 960 的

全局物件識別碼清單 (GOID list) 及全局粗糙特徵集 (GRFS)、來自方塊 958 的本地身分識別清單 (LIID list)、及來自方塊 956 的各類型的融合上下文區域及融合感興趣區域以進行全局細緻特徵對應處理以產生一全局細緻特徵集 (GDFS)。

【0188】 在方塊 964，進行全局細緻特徵融合 (GDFF)。舉例來說，全局細緻特徵融合模組 1324 係依據來自加權參數選擇模組 1323 所輸出的加權參數以執行全局細緻特徵融合處理，例如是將全局物件及特徵集對應模組 1321 所產生的全局細緻特徵集融合為一全局融合特徵。其中加權參數例如可為自適應加權係數 (AWC) 或鄰近可區分度加權係數 (ADWC)，視全局上下文分析之結果而定，其細節可參考第 8C-1 ~ 8C-2 圖及第 9B-1 ~ 9B-2 圖之實施例。

【0189】 在方塊 966，進行全局物件辨識 (GOR)。舉例來說，全局辨識模組 133 係將全局細緻特徵融合模組 1324 所產生的全局融合特徵輸入一全局物件辨識模型以辨識各全局融合特徵的全局身分，例如可建立記錄各全局融合特徵的全局身份識別碼的一全局身分識別清單。此外，全局身分識別清單更記錄了全局物件辨識模型的辨識結果及其信心度。

【0190】 在一些實施例中，例如在缺乏光線的黑暗環境中，可以透過耳朵的聽覺 (例如可利用麥克風 110B)，聽聞人的說話的聲調來判定講話的是誰。在一些實施例中，透過聽聞動物的叫聲來判定是狗叫、貓叫、或是其他動物的叫聲，甚至在經常接觸特定動物且熟稔其叫聲的狀況下，只靠聽聞叫聲就能判定是哪個特定的動物叫聲，聽出是某個鄰居的狗在叫。

【0191】 在一些實施例中，可透過氣味嗅覺(例如可利用氣味感測器 110D)來偵測周遭環境，並預測可能有危險，例如聞到燒焦味、聞到瓦斯味、聞到汽油味等。

【0192】 在一些實施例中，當在與他人對話或談判的情境下，透過聽聞對方的言語音調，觀察其行為舉止，甚至從對方身上發出的味道，例如有酒味或香水味等。

【0193】 監控系統 100 可融合上述來自不同類型之感測器(可類比於不同感官)所採集到的感測資訊，統合後再來做出適切的反應。詳細而言，監控系統 100 所監測的場景中可能會有部分的感測器之感測資料並沒有偵測到任何需要關注的物件，例如在黑暗環境或低光源的環境下，攝影機 110A 所擷取的視頻資料則通常無助於辨識全局物件。此時，運算裝置 120 可能會判斷出視頻資料並沒有值得關注的物件，但是可由音頻資料或其他類型的感測資料判斷出有值得關注的物件。此外，因為有部分類型的細緻特徵無助於辨識全局物件，故運算裝置 120 所執行的上下文分析處理的判斷結果會使用鄰近可區分度加權係數以進行全局細緻特徵融合處理。意即可將視頻資料有關視頻細緻特徵(例如包括顏色、紋理、及形狀)的加權係數均設定為 0，並進行全局細緻特徵融合處理。

【0194】 類似地，在另一實施例中，監控系統 100 所監測的場景可能為一吵雜環境，且麥克風 110B 所接收的音頻資料中可能混合了各種環境噪音，此時，雖然運算裝置 120 可判斷出音頻資料有值得關注的物件，但是判斷結果可能受到噪音的影響，使得判斷音頻物件的信心度(或準確度)下降。運算裝置 120

所執行的上下文分析處理的判斷結果會使用鄰近可區分度加權係數以進行全局細緻特徵融合處理。意即可將音頻資料有關音頻細緻特徵(例如包括音量、音高、及音色)的加權係數均設定為0，並進行全局細緻特徵融合處理。

【0195】第10圖係顯示依據本發明一實施例中之一場景及監控系統的示意圖。

【0196】傳統的視頻監視系統，每個攝影機是獨立持續地拍攝視頻影像並儲存在視頻監視系統的硬碟內。裝設攝影機後，視頻監視系統可將拍攝到的影像即時顯示在監視螢幕上，由安全人員隨時監看。若有事件發生，需要靠人工調閱硬碟內的影像檔案，但由於拍攝角度與設置空間位置等因素，還得靠人工去追蹤銜接不同攝影機所拍攝下來的影像畫面。因為各個獨立的攝影機之間的資訊不能即時融合交流，也缺乏跨類型感測融合分析的能力，容易受到光線干擾、遮蔽、物件交叉重疊等狀況，無法獲取完整資訊，導致辨識以偏概全且辨識結果不穩定。

【0197】本發明之監控系統可解決上述問題。如第10圖所示，場景1000例如為一銀行大門口附近的區域，其中銀行大門口上安裝有攝影機1010及一指向性麥克風1020，用於監看銀行大門1001之出入情況，且此區域例如定義為第一區域1031。由銀行大門1001進入銀行後，有一客戶等待區，例如設有沙發1002。且客戶等待區架設了攝影機1011及一指向性麥克風1021，用於監看由銀行大門進入到銀行大廳及客戶等待區的範圍，且此監看範圍係定義為第二區域1032。第一區域1031及第

二區域 1032 有一重疊區域，例如定義為第三區域 1033，且攝影機 1010～1011 及指向性麥克風 1020～1021 均能監看第三區域 1033。

【0198】在監控系統 100 中之各類型之感測器所安裝的位置與其所監控拍攝與收音的場景空間，能被運用世界座標定位資訊。所有被偵測辨識到的物件其所在空間位置能被轉換成世界座標定位。此外，所有的感測器所收集到的資訊可傳送至一中央資料處理中心(未繪示)，中央資料處理中心可以執行本發明前述實施例之方法以將各類型之感測器的感測資料經由 AI 辨識系統融合運用，並且產生回饋達到自我訓練自我強化偵測與辨識的能力與精確度。

【0199】舉例來說，若有三人(例如分別為人物 1041、1042、及 1043)到銀行大門並進入第一區域 1031 時，會被攝影機 1010 及指向性麥克風 1020 偵測到，且運算裝置 120 可利用上述實施例中之流程分別對人物 1041、1042、及 1043 建立了相應的標籤 ID#01、ID#02、及 ID#03。此外，運算裝置 120 更對人物 1041、1042、及 1043 被拍攝到的視頻資料及被收錄的音頻資料分別進行細緻特徵抽取，例如視頻細緻特徵包括：顏色、紋理、及形狀，且音頻細緻特徵包括：音量、音高、音色。

【0200】人物 1041 的標籤 ID#01 相應的視頻細緻特徵例如為：黑色、無條紋、大人、女性，且音頻細緻特徵例如為：很大聲、尖噪、明朗。人物 1042 的標籤 ID#02 相應的視頻細緻特徵例如為：藍色、無條紋、大人、男性，且音頻細緻特徵例如為：中等音量、渾厚深沉、飽滿。人物 1043 的標籤 ID#03 相應

的視頻細緻特徵例如為：黑色、橫條紋、小孩，且音頻細緻特徵例如為：大聲、明亮清晰、有活力。

【0201】請同時參考第 9A-1 及 9A-2 圖及第 10 圖，詳細而言，當人物 1041、1042、及 1043 位於第一區域 1031 時，運算裝置 120 係可擷取人物 1041、1042、及 1043 之標籤 ID#01、ID#02、及 ID#03 相應的視頻細緻特徵及音頻細緻特徵，並將標籤 ID#01、ID#02、及 ID#03 相應的視頻細緻特徵及音頻細緻特徵輸入個別的本地方物件辨識模型(例如方塊 906-1 及 906-2)，並產生個別的視頻物件及音頻物件的辨識結果，例如記錄於用於視頻物件的一本地身分識別清單 L1 及用於音頻物件的一本地身分識別清單 L2。

【0202】接著，經過方塊 908 及 910 以進行全局物件對應及全局細緻特徵集對應以產生本地物件識別碼清單、及全局粗糙特徵集及相應的全局細緻特徵集。假設在方塊 912 中是選擇自適應加權係數(AWC)更新，本地物件識別碼清單、及全局粗糙特徵集及相應的全局細緻特徵集會在方塊 918 中進行全局細緻特徵融合以產生全局融合特徵，並在方塊 920 的全局物件辨識模型中產生辨識結果，例如可辨識出全局物件 P1、P2、及 P3，並指派相應的全局身分識別碼 GIID1、GIID2、及 GIID3 至所辨識出的全局物件 P1、P2、及 P3。

【0203】簡單來說，全局身分識別碼 GIID1、GIID2、及 GIID3 所相應的全局物件即分別帶有上述標籤 ID#01、ID#02、及 ID#03 中的所有視頻細緻特徵及音頻細緻特徵。

【0204】因此，當人物 1041、1042、及 1043 進入第一區域

1031時，運算裝置120即已建立出人物1041、1042、及1043相應的ID#01、ID#02、及ID#03、以及其全局身分識別碼GIID1、GIID2、及GIID3與相應的所有視頻細緻特徵及音頻細緻特徵。

【0205】當人物1041、1042、及1043由第一區域1031進入第三區域1033時，因為第三區域1033已可同時由攝影機1010～1011及指向性麥克風1020～1021進行監看。在人物1041、1042、及1043由第一區域1031進入第三區域1033的過程中，攝影機1010及指向性麥克風1020會持續地收集視頻與音頻資料並進行物件追蹤與辨識，但可能會因為感測器設置的位置、角度與當時環境的光線背景聲音，亦或是人員交疊等遮蔽等狀況導致特徵資訊有所缺漏。因此，當人物1041、1042、及1043由第一區域1031進入第三區域1033後，除了可利用攝影機1010及指向性麥克風1020所擷取之視頻資料及音頻資料之外，運算裝置120可利用攝影機1011及指向性麥克風1021在不同位置及角度所擷取之視頻資料及音頻資料再依據上述步驟建立出各全局物件的全局融合特徵。此外，運算裝置120還能融合攝影機1010及指向性麥克風1020所收集到的特徵資料。再經過方塊920之全局物件辨識模型判斷由上述全局融合特徵是否與先前所辨識出的全局身分識別碼GIID1、GIID2、及GIID3所相應的全局物件的全局融合特徵相同。若全局融合特徵相同，則可判斷是相同的人物；若全局融合特徵不同，則可判斷是不同的人。

【0206】當人物1041、1042、及1043離開重疊區域1033並進入第二區域1032時，運算裝置120僅能利用攝影機1011及指

向性麥克風 1021 所擷取之視頻資料及音頻資料再依據上述步驟建立出各物件的全局融合特徵。再經過方塊 920 之全局物件辨識模型判斷由上述全局融合特徵是否與先前所辨識出的全局身分識別碼 GIID1、GIID2、及 GIID3 所相應的全局物件的全局融合特徵相同。若全局融合特徵相同，則運算裝置 120 可判斷是相同的人物；若全局融合特徵不同，則運算裝置 120 可判斷是不同的人。因此，本發明之監控系統用於進行物件辨識的資訊比傳統的監控系統更為豐富且完整，還有回饋補強的機制，可以讓追蹤與辨識的能力與精確度大幅提高。

【0207】第 11 圖係顯示依據本發明一實施例中之利用多維度感測器資料的監控方法之流程圖。請同時參考第 2 圖及第 11 圖。

【0208】在步驟 S1110，利用感測器（例如感測器 110A - 110E）偵測該場景以取得各類型的一感測資料。感測器 110 係包括多種不同類型的感測器，例如：攝影機 110A、麥克風 110B、味覺感測器 110C、氣味感測器 110D、觸覺感測器 110E、或其組合，但本發明之實施例並不限定於上述類型或屬性的感測器。

【0209】在步驟 S1120，分別對各類型的該感測資料進行一本地物件處理以產生一本地物件特徵資訊。舉例來說，本地物件處理例如包括本地物件偵測及對應模組 1311、一本地物件特徵抽取及融合模組 1312、及一本地物件辨識模型 1313 所執行的關於本地物件的各種處理。此外，本地物件辨識模組 131 所產生的本地物件特徵資訊包括各類型的感測資料之本地物件識

別碼清單、本地粗糙特徵集、本地融合特徵、及本地身分識別清單。

【0210】 在步驟 S1130，依據該本地物件特徵資訊進行一全局物件處理以產生一全局物件特徵資訊。舉例來說，全局物件處理例如包括全局物件及特徵集對應模組 1321、上下文區域分析模組 1322、加權參數選擇模組 1323、及全局細緻特徵融合模組 1324 所執行關於全局物件的各種處理。此外，特徵融合模組 132 所產生的全局物件特徵資訊包括：全局物件識別碼清單及相應的全局細緻特徵集、及全局融合特徵。

【0211】 在步驟 S1140，對該全局物件特徵資訊進行一全局物件辨識以產生一全局辨識結果。舉例來說，全局辨識模組 133 係將全局細緻特徵融合模組 1324 所產生的全局融合特徵輸入一全局物件辨識模型以辨識各全局融合特徵的全局身分，例如可建立記錄各全局融合特徵的全局身份識別碼的一全局身分識別清單。

【0212】 綜上所述，本發明之實施例係提供一種利用多維度感測器資料之監控系統及監控方法，其可利用不同類型之感測器以取得場景的感測資料，並進行同類型之本地物件的偵測、對應、及辨識，且可利用不同類型之本地物件進行對應以產生全局的感測資料之全局物件，其具有全局融合特徵。此外，本發明之實施例中之利用多維度感測器資料之監控系統及監控方法更可執行全局的全局物件辨識，使得監控場景中之物件的可靠度及準確度更高。

【0213】 本發明雖以較佳實施例揭露如上，然其並非用以

限定本發明的範圍，任何所屬技術領域中具有通常知識者，在不脫離本發明之精神和範圍內，當可做些許的更動與潤飾，因此本發明之保護範圍當視後附之申請專利範圍所界定者為準。

【符號說明】**【0214】**

- 100～監控系統；
- 110～感測器；
- 120～運算裝置；
- 121～儲存單元；
- 130～監控程式；
- 110A～攝影機；
- 110B～麥克風；
- 110C～味覺感測器；
- 110D～氣味感測器；
- 110E～觸覺感測器；
- 110A-1～110A-4～攝影機；
- 110B-1～110B-3～麥克風；
- O1、O2、O3～物件；
- 131～本地物件辨識模組
- 132～特徵融合模組；
- 133～全局辨識模組；
- 1311～本地物件偵測及對應模組；
- 1312～本地物件特徵抽取及融合模組；

1313～本地物件辨識模型；

1314～回饋路徑；

1321～全局物件及特徵集對應模組；

1322～上下文區域分析模組；

1323～加權參數選擇模組；

1331～回饋路徑；

1324～全局細緻特徵融合模組；

700～視頻幀；

710～感興趣區域；

715～視頻物件；

720～探索區域；

730～上下文區域；

750～音頻段；

755～音頻物件；

760～感興趣區域；

770～探索區域；

780～上下文區域；

S302－S324～步驟；

S402－S424、S502－S514～步驟；

S802－S812～步驟；

S8021－S8025、S8041－S8043～步驟；

S8201－S8217～步驟；

S832－S840、S850－S856～步驟；

TF1－TF7～時間幀；

ROI1a – ROI7a ~ 感興趣區域；

Context1 – Context7 ~ 上下文區域；

902-1 – 902-N ~ 方塊；

904-1 – 904-N ~ 方塊；

904-1A、904-1B、904-1C、904-1D、904-1E ~ 方塊；

904-2A、904-2B、904-2C、904-2D、904-2E ~ 方塊；

904-NA、904-NB、904-NC、904-ND、904-NE ~ 方塊；

906-1 – 906-N、912-1 – 912-N ~ 方塊；

9120、908、910、912、916、918、920、924、926 ~ 方塊；

914、922 ~ 選擇器；

952 – 966 ~ 方塊；

959、967 ~ 路徑；

Video1、Video2、Video3 ~ 視頻資料；

Audio1、Audio2、Audio3 ~ 音頻資料；

Smell1、Smell2、Smell3 ~ 氣味資料；

600、1000 ~ 場景；

1001 ~ 銀行大門；

1002 ~ 沙發；

1010 – 1011 ~ 攝影機；

1020 – 1021 ~ 指向性麥克風；

1041 – 1043 ~ 人物；

1031 ~ 第一區域；

1032 ~ 第二區域；

1033 ~ 重疊區域；

S1110 - S1140 ~ 步驟。

申請專利範圍

1. 一種利用多維度感測器資料的監控方法，用於一監控系統，該監控系統包括設置於一場景中的複數個感測器，且該等感測器係分為複數個類型，該監控方法包括：
利用該等感測器偵測該場景以取得各類型的一感測資料；
分別對各類型的該感測資料進行一本地物件處理以產生一本地物件特徵資訊；
依據該本地物件特徵資訊進行一全局物件處理以產生一全局物件特徵資訊；以及
對該全局物件特徵資訊進行一全局物件辨識以產生一全局辨識結果；
其中分別對各類型的該感測資料進行該本地物件處理以產生該本地物件特徵資訊之步驟包括：
分別對各類型的該感測資料進行一本地物件偵測及對應處理，以取得各類型的該感測資料之一本地物件識別碼清單及相應的一本地粗糙特徵集；
依據各類型的該感測資料所相應的該本地物件識別碼清單及相應的該本地粗糙特徵集，分別對各類型的該感測資料進行一本地物件特徵抽取及融合處理，以取得各類型的該感測資料之複數個本地細緻特徵集並將該等本地細緻特徵集融合以產生一本地融合特徵；以及
對各類型的該感測資料相應的該本地融合特徵分別輸入一本地物件辨識模型，以取得各類型的該感測資料的一本地身分識別清單。

2. 如申請專利範圍第1項所述之利用多維度感測器資料的監控方法，其中該等類型之該等感測器包括：複數個攝影機、複數個麥克風、複數個味覺感測器、複數個氣味感測器、複數個觸覺感測器、或其組合。
3. 如申請專利範圍第1項所述之利用多維度感測器資料的監控方法，其中各類型的該感測資料之該本地物件識別碼清單包括各類型的該感測資料中的一或多個本地物件，且各本地物件係具有相應的一本地物件識別碼；其中該本地粗糙特徵集包括各類型的該感測資料所相應的方向、距離、及概略。
4. 如申請專利範圍第3項所述之利用多維度感測器資料的監控方法，其中該本地物件特徵抽取及融合處理包括：依據各類型的該感測資料所相應的該本地物件識別碼清單及相應的該本地粗糙特徵集，抽取各類型的該感測資料之複數個本地細緻特徵以建立相應的一本地細緻特徵集；以及將各類型的該感測資料之該本地細緻特徵集融合為各本地物件之該本地融合特徵。
5. 如申請專利範圍第1項所述之利用多維度感測器資料的監控方法，其中該本地物件特徵資訊包括各類型的該感測資料之該本地物件識別碼清單、該本地粗糙特徵集、該本地融合特徵、及該本地身分識別清單，且依據各類型的該感測資料所相應的該本地物件特徵資訊，進行該全局物件處理以產生該全局物件特徵資訊之步驟包括：

依據該本地物件特徵資訊進行一全局物件對應處理，以產生一全局物件識別碼清單及相應的一全局粗糙特徵集；以及

依據該本地物件特徵資訊、該全局物件識別碼清單及相應的該全局粗糙特徵集，進行一全局特徵集對應處理以產生在該全局物件識別碼清單中之一或多個全局物件之每一者相應的一全局細緻特徵集。

6. 如申請專利範圍第5項所述之利用多維度感測器資料的監控方法，其中對該全局物件特徵資訊進行一全局物件辨識以產生該全局辨識結果之步驟包括：

對各類型的該感測資料進行一本地上下文分析處理，以產生一本地上下文分析結果，並合併各本地上下文分析結果以產生一本地上下文合併結果；

依據該本地上下文合併結果以選擇一鄰近可區分度加權係數或一自適應加權係數；以及

依據所選擇的該鄰近可區分度加權係數或該自適應加權係數，進行一全局細緻特徵融合處理以產生各全局物件相應的一全局融合特徵。

7. 如申請專利範圍第6項所述之利用多維度感測器資料的監控方法，其中該全局辨識結果包括一信心度，且該方法更包括：

將該全局融合特徵拆解為各類型之該感測資料相應的複數個本地細緻特徵；以及

將該信心度及各類型之該感測資料相應的該等本地細緻特

徵分別回饋至各類型之該感測資料相應的該本地物件辨識模型。

8. 如申請專利範圍第6項所述之利用多維度感測器資料的監控方法，其中該本地上下文分析處理包括：

對各類型之該感測資料進行一上下文獲取處理以取得相應的一上下文區域；

對各類型之該感測資料相應的各上下文區域進行一上下文融合處理，以得到各類型之該感測資料相應的一融合上下文區域；

對各類型之該感測資料相應的該融合上下文區域分別進行該本地上下文分析處理，以產生該本地上下文分析結果；

以及

合併各本地上下文分析結果以產生該本地上下文合併結果。

9. 如申請專利範圍第1項所述之利用多維度感測器資料的監控方法，更包括：

當在不同類型之該感測資料中之複數個本地物件的一時間戳記、該本地粗糙特徵集、及世界座標定位資訊均吻合時，

判斷該等本地物件係對應成功；以及

指派一全局物件識別碼至對應成功的該等本地物件。

10. 一種監控系統，包括：

複數個感測器，其中該等感測器係分為複數個類型，且用以偵測一場景以取得各類型的一感測資料；以及

一運算裝置，用以分別對各類型的該感測資料進行一本地

物件處理，以產生一本地物件特徵資訊；

其中該運算裝置更依據該本地物件特徵資訊進行一全局物件處理以產生一全局物件特徵資訊，並對該全局物件特徵資訊進行一全局物件辨識以產生一全局辨識結果；

其中該運算裝置係分別對各類型的該感測資料進行一本地物件偵測及對應處理，以取得各類型的該感測資料之一本地物件識別碼清單及相應的一本地粗糙特徵集；

其中該運算裝置係依據各類型的該感測資料所相應的該本地物件識別碼清單及相應的該本地粗糙特徵集，分別對各類型的該感測資料進行一本地物件特徵抽取及融合處理，以取得各類型的該感測資料之複數個本地細緻特徵集並將該等本地細緻特徵集融合以產生一本地融合特徵；

其中該運算裝置係對各類型的該感測資料相應的該本地融合特徵分別輸入一本地物件辨識模型，以取得各類型的該感測資料的一本地身分識別清單。

11. 如申請專利範圍第10項所述之監控系統，其中該等類型之該等感測器包括：複數個攝影機、複數個麥克風、複數個味覺感測器、複數個氣味感測器、複數個觸覺感測器、或其組合。

12. 如申請專利範圍第10項所述之監控系統，其中各類型的該感測資料之該本地物件識別碼清單包括各類型的該感測資料中的一或多個本地物件，且各本地物件係具有相應的一本地物件識別碼；

其中該本地粗糙特徵集包括各類型的該感測資料所相應的

方向、距離、及概略。

13. 如申請專利範圍第12項所述之監控系統，其中該運算裝置係依據各類型的該感測資料所相應的該本地物件識別碼清單及相應的該本地粗糙特徵集，抽取各類型的該感測資料之複數個本地細緻特徵以建立相應的一本地細緻特徵集，並將各類型的該感測資料之該本地細緻特徵集融合為各本地物件之該本地融合特徵。
14. 如申請專利範圍第10項所述之監控系統，其中該本地物件特徵資訊包括各類型的該感測資料之該本地物件識別碼清單、該本地粗糙特徵集、該本地融合特徵、及該本地身分識別清單，且該運算裝置更依據該本地物件特徵資訊進行一全局物件對應處理，以產生一全局物件識別碼清單及相應的一全局粗糙特徵集，並依據該本地物件特徵資訊、該全局物件識別碼清單及相應的該全局粗糙特徵集，進行一全局特徵集建立處理以產生在該全局物件識別碼清單中之一或多個全局物件之每一者相應的一全局細緻特徵集。
15. 如申請專利範圍第14項所述之監控系統，其中該運算裝置更對各類型的該感測資料進行一上下文分析處理以產生一上下文分析結果，並合併各上下文分析結果以產生一上下文合併結果，並依據該上下文合併結果以選擇一鄰近可區分度加權係數或一自適應加權係數；
其中該運算裝置更依據所選擇的該鄰近可區分度加權係數或該自適應加權係數進行一全局細緻特徵融合處理以

產生各全局物件相應的一全局融合特徵。

16. 如申請專利範圍第15項所述之監控系統，其中該全局辨識結果包括一信心度，且該運算裝置更將該全局融合特徵拆解為各類型之該感測資料相應的複數個本地細緻特徵，並將該信心度及各類型之該感測資料相應的該等本地細緻特徵分別回饋至各類型之該感測資料相應的該本地物件辨識模型。
17. 如申請專利範圍第15項所述之監控系統，其中該運算裝置更對各類型之該感測資料進行一上下文獲取處理以取得相應的一上下文區域，並對各類型之該感測資料相應的各上下文區域進行一上下文融合處理以得到各類型之該感測資料相應的一融合上下文區域；
其中該運算裝置更對各類型之該感測資料相應的該融合上下文區域分別進行該上下文分析處理，以產生該上下文分析結果，並合併各上下文分析結果以產生該上下文合併結果。
18. 如申請專利範圍第10項所述之監控系統，其中當在不同類型之該感測資料中之複數個本地物件的一時間戳記、該本地粗糙特徵集、及世界座標定位資訊均吻合時，該運算裝置係判斷該等本地物件係對應成功，並指派一全局物件識別碼至對應成功的該等本地物件。