



(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 I759286 B

(45)公告日：中華民國 111 (2022) 年 04 月 01 日

(21)申請案號：106108280

(22)申請日：中華民國 106 (2017) 年 03 月 14 日

(51)Int. Cl. : G06F16/00 (2019.01)

G06T7/00 (2017.01)

(30)優先權：2016/03/17 美國

62/309,777

(71)申請人：加拿大商艾維吉隆股份有限公司 (加拿大) AVIGILON CORPORATION (CA)
加拿大

(72)發明人：雪利凡斯塔瓦 艾希許 SHRIVASTAVA, ASHISH (IN)

(74)代理人：陳長文

(56)參考文獻：

TW I340930

TW I442326

US 2009/0161962A1

US 2015/0054824A1

US 2016/0070986A1

審查人員：高元良

申請專利範圍項數：17 項 圖式數：9 共 58 頁

(54)名稱

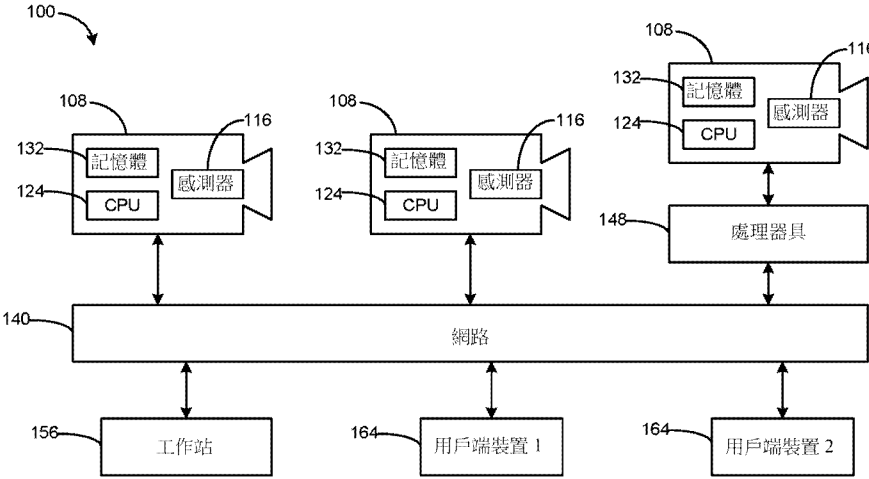
用於藉由機器學習訓練物件分類器之系統及方法

(57)摘要

一種用於訓練一電腦實施之物件分類器之系統及方法包含：偵測一場景之一子區域內之一前景視覺物件；判定該場景之該子區域之一背景模型，該背景模型表示在該子區域中不存在任何前景視覺物件時之彼子區域；及藉由使用該子區域之該背景模型作為一負向訓練實例進行電腦實施之機器學習來訓練該物件分類器。

A system and method for training a computer-implemented object classifier includes detecting a foreground visual object within a sub-region of a scene, determining a background model of the sub-region of the scene, the background model representing the sub-region when any foreground visual object is absent from that sub-region, and training the object classifier by computer-implemented machine learning using the background model of the sub-region as a negative training example.

指定代表圖：



【圖 1A】

符號簡單說明：

- 100 . . . 視訊擷取與播放系統/影像擷取與播放系統/系統
- 108 . . . 視訊擷取裝置/影像擷取裝置
- 116 . . . 影像感測器/感測器
- 124 . . . 處理器
- 132 . . . 記憶體裝置
- 140 . . . 網路
- 148 . . . 處理器具
- 156 . . . 工作站
- 164 . . . 用戶端裝置



【中文發明名稱】

用於藉由機器學習訓練物件分類器之系統及方法

【英文發明名稱】

SYSTEM AND METHOD FOR TRAINING OBJECT CLASSIFIER
BY MACHINE LEARNING

【中文】

一種用於訓練一電腦實施之物件分類器之系統及方法包含：偵測一場景之一子區域內之一前景視覺物件；判定該場景之該子區域之一背景模型，該背景模型表示在該子區域中不存在任何前景視覺物件時之彼子區域；及藉由使用該子區域之該背景模型作為一負向訓練實例進行電腦實施之機器學習來訓練該物件分類器。

【英文】

A system and method for training a computer-implemented object classifier includes detecting a foreground visual object within a sub-region of a scene, determining a background model of the sub-region of the scene, the background model representing the sub-region when any foreground visual object is absent from that sub-region, and training the object classifier by computer-implemented machine learning using the background model of the sub-region as a negative training example.

【指定代表圖】

圖1A

【代表圖之符號簡單說明】

100	視訊擷取與播放系統/影像擷取與播放系統/系統
108	視訊擷取裝置/影像擷取裝置
116	影像感測器/感測器
124	處理器
132	記憶體裝置
140	網路
148	處理器具
156	工作站
164	用戶端裝置

【發明說明書】

【中文發明名稱】

用於藉由機器學習訓練物件分類器之系統及方法

【英文發明名稱】

SYSTEM AND METHOD FOR TRAINING OBJECT CLASSIFIER
BY MACHINE LEARNING

【技術領域】

本發明標的物係關於視覺物件之分類，且更特定而言係關於使用所偵測前景視覺物件之背景模型作為負向訓練實例來訓練一電腦實施之物件分類器。

【先前技術】

電腦實施之視覺物件分類(亦稱作物件辨識)涉及將存在於由一攝影機擷取之靜態影像或運動視訊中之現實物件之視覺表示分類。藉由執行視覺物件分類，根據存在於靜態影像或運動視訊中之每一視覺物件之類型(例如人類、交通工具、動物)而將其分類。

自動保全與監控系統通常採用視訊攝影機或者其他影像擷取裝置或感測器來收集影像資料。在最簡單系統中，由影像資料表示之影像經顯示以用於由保全人員同時篩選及/或經記錄以用於稍後在一保全違反之後參考。在彼等系統中，對所關注視覺物件進行偵測及分類之任務由一人類觀察者執行。當系統本身能夠部分地或完全地執行物件偵測及分類時，發生一顯著進步。

舉例而言，在一典型監控系統中，人們可能關注偵測在環境中移動之諸如人類、交通工具、動物等物件。不同物件可能構成不同威脅或警報

級別。舉例而言，場景中之一動物可係正常的，但場景中之一人類或交通工具可造成一警報且可需要一保全警衛之立即關注。對由攝影機所擷取之影像資料表示之影像中之物件之自動電腦實施之偵測及分類可顯著促進保全人員之篩選任務以及改良影像資料之記錄。

【發明內容】

本文中所闡述之實施例在一項態樣中提供一種用於訓練一電腦實施之物件分類器之方法。該方法包含：偵測一場景之一子區域內之一前景視覺物件；判定該場景之該子區域之一背景模型，該背景模型表示在其中不存在任何前景視覺物件時之該子區域；及藉由使用該子區域之該背景模型作為一負向訓練實例進行電腦實施之機器學習來訓練該物件分類器。

本文中所闡述之實施例在另一態樣中提供一種電腦實施之物件分類器。系統包含：一處理器；一電腦可讀儲存裝置，其儲存在由該處理器執行時致使該系統執行包含以下各項之操作之程式指令：偵測一場景之一子區域內之一前景視覺物件；判定該場景之該子區域之一背景模型，該背景模型表示在其中不存在任何前景視覺物件時之該子區域；及藉由使用該子區域之該背景模型作為一負向訓練實例進行電腦實施之機器學習來訓練該物件分類器。

根據某些實例性實施例，該等方法及/或系統進一步包含藉由使用該所偵測前景視覺物件作為一正向訓練實例進行機器學習來訓練該物件分類器。

根據某些實例性實施例，判定該場景之該子區域之該背景模型包含：選擇在對應於該場景之該子區域的一歷史影像圖框之一子區域中不存在任何前景物件時擷取之該歷史影像圖框；及自該歷史影像圖框裁剪對應

於該場景之該子區域之該子區域，該經裁剪影像圖框係該場景之該子區域之該背景模型。

根據某些實例性實施例，判定該場景之該子區域之該背景模型包含：在複數個歷史影像圖框中之每一者內判定不含任何前景物件之一或多個子區域；聚合來自該複數個歷史影像圖框之該一或多個子區域以形成表示該整個場景之一完整背景影像；及自該完整背景影像裁剪對應於該場景之該子區域之一子區域，該經裁剪完整背景影像係該場景之該子區域之該背景模型。

根據某些實例性實施例，聚合來自該複數個歷史影像圖框之該一或多個子區域包括拼接該一或多個子區域以形成表示該整體場景之一影像。

根據某些實例性實施例，專門針對一當前場景而訓練該物件分類器。

根據某些實例性實施例，在該當前場景被改變為一新場景時，恢復為未進行該當前場景所特有之該訓練之該物件分類器；且藉由使用來自該新場景之背景模型之機器學習來訓練該物件分類器。

根據某些實例性實施例，部分地使用受監督學習來準備該物件分類器。

根據某些實例性實施例，該電腦實施之機器學習選自迴旋神經網路、支援向量機器、決策樹、隨機森林及級聯分類器。

根據某些實例性實施例，該等方法及/或系統進一步包含藉由使用一場景之一經錯誤分類子區域作為一負向訓練實例之電腦實施之機器學習來訓練該物件分類器。

【圖式簡單說明】

具體實施方式參考以下各圖，其中：

圖1A圖解說明根據一實例性實施例之一視訊擷取與播放系統之經連接裝置之一方塊圖；

圖1B圖解說明根據一項實例性實施例之視訊擷取與播放系統之一操作模組集合之一方塊圖；

圖1C圖解說明根據一項實例性實施例之在一個裝置內實施之一操作模組集合之一方塊圖；

圖2圖解說明用於對影像資料執行視訊分析之一方法之一實例性實施例之一流程圖；

圖3A圖解說明根據一項實例性實施例之一視訊分析模組之一操作子模組集合之一方塊圖；

圖3B圖解說明根據一項實例性實施例之一物件分類模組之複數個物件分類器；

圖4圖解說明此項技術中已知之用於一基礎分類器之進一步訓練之一方法之一流程圖；

圖5圖解說明根據一項實例性實施例之用於一基礎分類器之進一步訓練之一經改良電腦實施之方法之一流程圖；

圖6A至圖6F係具有所偵測前景視覺物件的場景之子區域及其對應背景模型；

圖7A係表示一場景之一實例之一第一全歷史影像圖框；

圖7B係表示場景之一實例之一第二全歷史影像圖框；

圖8圖解說明根據一替代實例性實施例之用於一基礎分類器之進一步訓練之一經改良電腦實施之方法之一流程圖；及

圖9圖解說明根據一項實例性實施例之用於一基礎分類器之場景特定訓練之一經改良電腦實施之方法之一流程圖。

將瞭解,為圖解說明之簡潔及清晰起見,未必按比例繪製圖中所展示之元件。舉例而言,為清晰起見,可相對於其他元件誇大某些元件之尺寸。此外,在認為適當之處,可在該等圖當中重複若干元件符號來指示對應或類似元件。

【實施方式】

敘述眾多特定細節以便提供對本文中所闡述之例示性實施例之一透徹理解。然而,熟習此項技術者將理解,可在不具有此等特定細節之情況下實踐本文中所闡述之實施例。在其他例項中,未詳細闡述眾所周知之方法、程序及組件以免使本文中所闡述之實施例模糊。此外,此說明不應以任何方式被視為限制本文中所闡述之實施例之範疇,而是應被視為僅闡述本文中所闡述之各項實施例之實施方案。

諸如「頂部」、「底部」、「向上」、「向下」、「垂直」及「橫向」等方向性術語在以下說明中係僅出於提供相對參考之目的而使用,且並非意欲暗示對任何物品在使用期間如何定位或在一總成中或相對於一環境如何安裝之任何限制。

除非另外明確規定,否則術語「一態樣」、「一實施例」、「實施例」、「若干實施例」、「該實施例」、「該等實施例」、「一或多項實施例」、「某些實施例」、「特定實施例」、「一項實施例」、「另一實施例」及類似術語意指「所揭示發明之一或多項(但非全部)實施例」。除非另外明確規定,否則在闡述一實施例時對「另一實施例」或「另一態樣」之一參考不隱含所參考實施例與另一實施例(例如,在所參考實施例

之前闡述之一實施例)係相互排斥的。

除非另外明確規定，否則術語「包含(including)」、「包括(comprising)」及其變化形式意指「包含但不限於」。

除非另外明確規定，否則術語「複數個」意指「兩個或多於兩個」。除非另外明確規定，否則術語「本文中」意指「在包含可以引用方式併入之任何事物之本申請案中」。

術語「例如(e.g.)」及類似術語意指「舉例而言(for example)」，且因此不限制其解釋之術語或片語。

術語「各別(respective)」及類似術語意指「個別來看(taken individually)」。因此，若兩個或多於兩個事物具有「各別」特性，那麼每一此事物具有其自身特性，且此等特性可彼此不同但不必彼此不同。舉例而言，片語「兩個機器中之每一者具有一各別功能」意味著第一此機器具有一功能且第二此機器亦具有一功能。第一機器之功能可或可不與第二機器之功能相同。

除非內容另外清楚指示，否則在申請專利範圍及/或說明書中字詞「一(a或an)」在聯合術語「包括」或「包含」一起使用時可意指「一個」，但其亦與「一或多個」、「至少一個」及「一個或多於一個」之含義一致。類似地，除非內容另外清楚指示，否則字詞「另一」可意指至少一第二或更多。

如本文中所使用之術語「經耦合(coupled)」、「耦合(coupling)」或「經連接(connected)」可取決於其中使用此等術語之內容脈絡而具有數個不同含義。舉例而言，術語經耦合、耦合或經連接可具有一機械或電內涵。舉例而言，如本文中所使用，術語經耦合、耦合或經連接可指示兩個

元件或裝置直接彼此連接或取決於特定內容脈絡透過一或多個中間元件或裝置經由一電元件、電信號或一機械元件而彼此連接。

本文中之「影像資料」係指由視訊擷取裝置產生且表示由視訊擷取裝置擷取之影像之資料。影像資料可包含一起形成由視訊擷取裝置擷取之一視訊之複數個順序影像圖框。每一影像圖框可由一像素矩陣表示，每一像素具有一像素影像值。舉例而言，像素影像值可係一灰階數值(例如，0至255)或針對彩色影像之複數個數值。用於表示影像資料中之像素影像值之色彩空間之實例包含RGB影像、YUV影像、CYKM影像、YCBCR 4:2:2影像及YCBCR 4:2:0影像。將理解，如本文中所使用之「影像資料」可係指由視訊擷取裝置產生之「原始」影像資料及/或已經歷某種形式處理之影像資料。

一「前景視覺物件」係指存在於由視訊擷取裝置擷取之影像圖框中之一現實物件(例如，人、動物、交通工具)之一視覺表示。前景視覺物件係出於各種目的(諸如視訊監控)而受關注之物件。舉例而言，一場景中之前景視覺物件可表示一事件，諸如所呈現之一人類或交通工具。一前景視覺物件可係一移動物件或一先前移動物件。前景視覺物件與一背景物件有區別，背景物件係存在於一場景之背景中之一物件且不受關注。

一「當前影像圖框」係指一視訊之複數個順序影像圖框內當前在本文中所闡述之各種系統及方法內分析之一影像圖框。一當前影像圖框之影像資料經分析以產生關於在當前影像圖框內及/或在當前影像之前的複數個影像圖框內擷取之物件之資訊。

一當前影像圖框之一「先前影像圖框」或一「歷史影像圖框」係指一視訊之複數個順序影像圖框內在一當前影像圖框之前發生之一影像圖

框。舉例而言，先前影像圖框可係緊鄰當前影像圖框之前的影像圖框。另一選擇係，先前影像圖框可係複數個順序影像圖框中之一較早影像圖框，但其足夠接近於當前影像圖框以便與當前影像圖框相關。

本文中之「處理影像資料」或其變體係指對影像資料執行之一或多個電腦實施之功能。舉例而言，處理影像資料可包含但不限於影像處理操作，對視訊資料進行分析、管理、壓縮、編碼、儲存、傳輸及/或播放。分析影像資料可包含將影像圖框之區分段及偵測視覺物件、對位於由影像資料表示之所擷取場景內之視覺物件進行追蹤及/或分類。處理影像資料可致使產生經修改影像資料，諸如經壓縮(例如品質降低)及/或經重新編碼影像資料。處理影像資料亦可致使輸出關於影像資料或在影像內擷取之視覺物件之額外資訊。舉例而言，此額外資訊通常理解為後設資料。後設資料亦可用於對影像資料之進一步處理，諸如圍繞影像圖框中之所偵測視覺物件畫定界框。

在兩個或多於兩個術語或片語同義之情況下(例如，由於該等術語或片語同義之一明確陳述)，一個此種術語/片語之例項並不意味著另一此種術語/片語之例項必須具有一不同含義。舉例而言，在一陳述提出「包含」之含義與「包含但不限於」同義之情況下，僅使用片語「包含但不限於」並不意味著術語「包含」意指除「包含但不限於」之外的事物。

標題(在本申請案之第一頁之開始處敘述)或摘要(在本申請案之開始處敘述)皆不應以任何方式被視為限制所揭示發明之範疇。本申請案中包含摘要僅僅係因為依據37 C.F.R. 1.72(b)節或其他司法管轄區中之類似法律要求不超過150字詞之一摘要。本申請案之標題及本申請案中所提供之章節之標頭僅係出於簡便起見，且不應以任何方式被視為限制本發明。

本申請案中闡述且僅出於圖解說明之目的而呈現眾多實施例。所闡述實施例在任何意義下並不且並不意欲具限制性。當前所揭示態樣廣泛適用於眾多實施例，如自本發明容易地明瞭。熟習此項技術者將認識到，所揭示態樣可在具有各種修改及更改(諸如結構及邏輯修改)之情況下實踐。雖然可參考一或多個特定實施例及/或圖式闡述所揭示態樣之特定特徵，但應理解，除非另外明確規定，否則此等特徵不限於在闡述其所參考之一或多個特定實施例或圖式中使用。

本申請案中所闡述之方法步驟或產品元件之實施例皆並非必要的或並非共同延伸的，本說明書中明確如此陳述或一技術方案中明確如此敘述之情況除外。

如熟習此項技術者將瞭解，本文中所闡述之各項實例性實施例可體現為一方法、系統或電腦程式產品。因此，各項實例性實施例可採取以下形式：一完全硬體實施例、一完全軟體實施例(包含韌體、常駐軟體、微碼等)或組合軟體與硬體態樣之一實施例，本文中可將所有該等軟體及硬體態樣統稱為一「電路」、「模組」或「系統」。此外，各項實例性實施例可採取一電腦可用儲存媒體上之一電腦程式產品之形式，該電腦可用儲存媒體具有體現在該媒體中之電腦可用程式碼。

可利用任何適合電腦可用或電腦可讀媒體。電腦可用或電腦可讀媒體可係(舉例而言但不限於)一電子媒體、磁性媒體、光學媒體、電磁媒體、紅外線媒體或者半導體系統、設備、裝置或傳播媒體。在本文件之內容脈絡中，一電腦可用或電腦可讀媒體可係可含有、儲存、傳遞、傳播或輸送程式以供指令執行系統、設備或裝置使用或者結合指令執行系統、設備或裝置使用之任何媒體。

用於執行各項實例性實施例之操作之電腦程式碼可以一物件導向程式設計語言(諸如Java、Smalltalk、C++、Python或類似語言)撰寫。然而，用於執行各項實例性實施例之操作之電腦程式碼亦可以習用程序程式設計語言(諸如「C」程式設計語言或類似程式設計語言)撰寫。程式碼可在一電腦上完全執行、在該電腦上作為一獨立軟體封裝部分執行、在該電腦上部分且在一遠端電腦上部分執行或在該遠端電腦或伺服器上完全執行。在後一情境中，該遠端電腦可透過一區域網路(LAN)或一廣域網路(WAN)連接至該電腦，或者可進行至一外部電腦之連接(舉例而言，透過使用一網際網路服務提供者之網際網路)。

下文參考根據本發明之實施例之方法、設備(系統)及電腦程式產品之流程圖圖解說明及/或方塊圖來闡述各項實例性實施例。將理解，該等流程圖圖解說明及/或方塊圖中之每一方塊及該等流程圖圖解說明及/或方塊圖中之方塊之組合皆可由電腦程式指令來實施。此等電腦程式指令可提供至一通用電腦、專用電腦或其他可程式化資料處理設備之一處理器以產生一機器，使得經由電腦或其他可程式化資料處理設備之處理器而執行之該等指令建立用於實施流程圖及/或方塊圖之一或若干方塊中所規定之功能/動作之構件。

此等電腦程式指令亦可儲存於一電腦可讀媒體中，該電腦可讀媒體可引導一電腦或其他可程式化資料處理設備以一特定方式起作用，使得儲存於電腦可讀記憶體中之指令產生包含實施流程圖及/或方塊圖之一或若干方塊中所規定之功能/動作之指令之一製品。

亦可將電腦程式指令載入至一電腦或其他可程式化資料處理設備上以致使在該電腦或其他可程式化設備上執行一系列操作步驟，以產生一電

腦實施處理程序，使得在該電腦或其他可程式化設備上執行之指令提供用於實施流程圖及/或方塊圖之一或若干方塊中所規定之功能/動作之步驟。

現在參考圖1A，其中圖解說明根據一實例性實施例之視訊擷取與播放系統100之經連接裝置之一方塊圖。舉例而言，視訊擷取與播放系統100可用作一視訊監控系統。視訊擷取與播放系統100包含執行本文中所闡述之程序及功能之硬體及軟體。

視訊擷取與播放系統100包含可操作以擷取複數個影像且產生表示複數個所擷取影像之影像資料之至少一個視訊擷取裝置108。

每一視訊擷取裝置108包含用於擷取複數個影像之至少一個影像感測器116。視訊擷取裝置108可係一數位視訊攝影機，且影像感測器116可輸出所擷取光作為一數位資料。舉例而言，影像感測器116可係一CMOS、NMOS或CCD。

至少一個影像感測器116可操作以擷取在一或多個頻率範圍中之光。舉例而言，至少一個影像感測器116可操作以擷取在實質上對應於可見光頻率範圍之一範圍中之光。在其他實例中，至少一個影像感測器116可操作以擷取在可見光範圍之外(諸如在紅外光及/或紫外光範圍中)的光。在其他實例中，視訊擷取裝置108可係包含可操作以擷取在不同頻率範圍中之光之兩個或多於兩個感測器之一多感測器攝影機。

至少一個視訊擷取裝置108可包含一專用攝影機。將理解，本文中專用攝影機係指其主要特徵係擷取影像或視訊之一攝影機。在某些實例性實施例中，專用攝影機可執行關聯於所擷取影像或視訊之功能，諸如但不限於處理由其或由另一視訊擷取裝置108產生之影像資料。舉例而言，專用攝影機可係一監控攝影機，諸如一轉動-傾斜-變焦攝影機、半球型攝影

機、天花板中攝影機、箱式攝影機及子彈型攝影機中之任一者。

另外或另一選擇係，至少一個視訊擷取裝置108可包含一嵌入式攝影機。將理解，本文中嵌入式攝影機係指嵌入於操作以執行與所擷取影像或視訊不相關之功能之一裝置內之一攝影機。舉例而言，嵌入式攝影機可係存在於一膝上型電腦、平板電腦、無人機裝置、智慧型電話、視訊遊戲控制台或控制器中之任一者上之一攝影機。

每一視訊擷取裝置108包含一或多個處理器124、耦合至處理器之一或多個記憶體裝置132及一或多個網路介面。記憶體裝置可包含在程式指令之執行期間採用之一本地記憶體(例如，一隨機存取記憶體及一快取記憶體)。處理器執行可儲存於記憶體裝置中之電腦程式指令(例如，一作業系統及/或應用程式)。

在各項實施例中，處理器124可由具有獨立地或並行地操作(包含可能冗餘地操作)之一或多個電路單元(包含數位信號處理器(DSP)、圖形處理單元(GPU)嵌入式處理器等及其任何組合)之任何處理電路實施。此處理電路可由一或多個積體電路(IC)實施，包含由一單片積體電路(MIC)、一特殊應用積體電路(ASIC)、一場可程式化閘陣列(FPGA)等或其任何組合實施。另外或另一選擇係，舉例而言，此處理電路可實施為一可程式化邏輯控制器(PLC)。舉例而言，處理器可包含用於儲存記憶體(諸如數位資料)之電路且可包括記憶體電路或與記憶體電路進行有線通信。

在各項實施例性實施例中，耦合至處理器電路之記憶體裝置132可操作以儲存資料及電腦程式指令。通常，記憶體裝置係一數位電子積體電路之全部或部分或者由複數個數位電子積體電路形成。舉例而言，記憶體裝置可實施為唯讀記憶體(ROM)、可程式化唯讀記憶體(PROM)、可抹除可程

式化唯讀記憶體(EPROM)、電可抹除可程式化唯讀記憶體(EEPROM)、快閃記憶體、一或多個快閃磁碟機、通用串列匯流排(USB)連接式記憶體單元、磁性儲存裝置、光學儲存裝置、磁光儲存裝置等或其任何組合。記憶體裝置可操作以儲存作為揮發性記憶體、非揮發性記憶體、動態記憶體等或其任何組合之記憶體。

在各項實例性實施例中，影像擷取裝置108之複數個該等組件可一起整合於一系統單晶片(SOC)內。舉例而言，處理器124、記憶體裝置132及網路介面可在一SOC內實施。此外，當以此方式實施時，一通用處理器以及一GPU及一DSP中之一或多者可一起在SOC內實施。

繼續圖1A，至少一個視訊擷取裝置108中之每一者連接至一網路140。每一視訊擷取裝置108可操作以輸出表示其擷取之影像之影像資料且經由網路傳輸影像資料。

將理解，網路140可係提供資料之接收及傳輸之任何通信網路。舉例而言，網路140可係一區域網路、外部網路(例如，WAN、網際網路)或其中之一組合。在其他實例中，網路140可包含一雲端網路。

在某些實例中，視訊擷取與播放系統100包含一處理器具148。處理器具148可操作以處理由一視訊擷取裝置108輸出之影像資料。處理器具148亦包含一或多個處理器及耦合至一處理器之一或多個記憶體裝置。處理器具148亦可包含一或多個網路介面。

舉例而言且如所圖解說明，處理器具148連接至一視訊擷取裝置108。處理器具148可進一步連接至網路140。

根據一項例示性實施例且如圖1A中所圖解說明，視訊擷取與播放系統100包含至少一個工作站156(例如，伺服器)，每一工作站具有一或多

個處理器。至少一個工作站156亦可包含儲存記憶體。工作站156自至少一個視訊擷取裝置108接收影像資料且執行對影像資料之處理。工作站156可進一步發送用於管理及/或控制影像擷取裝置108中之一或多者之命令。工作站156可自視訊擷取裝置108接收原始影像資料。另一選擇係或另外，工作站156可接收已經歷某種中間處理(諸如視訊擷取裝置108處及/或一處理器具148處之處理)之影像資料。工作站156亦可接收來自影像資料之後設資料且執行對影像資料之進一步處理。

將理解，儘管圖1A中圖解說明一單個工作站156，但該工作站可實施為複數個工作站之一聚合。

視訊擷取與播放系統100進一步包含連接至網路140之至少一個用戶端裝置164。用戶端裝置164由一或多個使用者用於與視訊擷取與播放系統100互動。因此，用戶端裝置164包含至少一個顯示裝置及至少一個使用者輸入裝置(例如，滑鼠、鍵盤、觸控螢幕)。用戶端裝置164可操作以在其顯示裝置上顯示用於顯示資訊、接收使用者輸入及播放視訊之一使用者介面。舉例而言，該用戶端裝置可係一個人電腦、膝上型電腦、平板電腦、個人資料助理(PDA)、行動電話、智慧型電話、遊戲裝置及其他行動裝置中之任一者。

用戶端裝置164可操作以經由網路140接收影像資料且進一步可操作以播放所接收影像資料。一用戶端裝置164亦可具有用於處理影像資料之功能性。舉例而言，一用戶端裝置164之處理功能可限於與播放所接收影像資料之能力相關之處理。在其他實例中，影像處理功能性可在工作站156與一或多個用戶端裝置164之間共用。

在某些實例中，影像擷取與播放系統100可在無工作站156之情況下

實施。因此，影像處理功能性可完全在一或多個視訊擷取裝置108上執行。另一選擇係，影像處理功能性可在視訊擷取裝置108、處理器具148及用戶端裝置164中之兩者或多於兩者當中共用。

現在參考圖1B，其中圖解說明根據一項實例性實施例之視訊擷取與播放系統100之一操作模組集合200的一方塊圖。該等操作模組可以硬體、軟體或兩者實施於如圖1A中所圖解說明之視訊擷取與播放系統100之裝置中之一或多者上。

操作模組集合200包含至少一個視訊擷取模組208。舉例而言，每一視訊擷取裝置108可實施一視訊擷取模組208。視訊擷取模組208可操作以控制一視訊擷取裝置108之一或多個組件(例如，感測器116等)以擷取影像。

操作模組集合200包含一影像資料處理模組子集216。舉例而言且如所圖解說明，影像資料處理模組子集216包含一視訊分析模組224及一視訊管理模組232。

視訊分析模組224接收影像資料且分析影像資料以判定所擷取影像或視訊及/或存在於由影像或視訊表示之場景中之物件之性質或特性。基於所做出之判定，視訊分析模組224可進一步輸出提供關於該等判定之資訊之後設資料。由視訊分析模組224做出之判定之實例可包含以下各項中之一或多者：前景/背景分段、物件偵測、物件追蹤、物件分類、虛擬絆網、異常偵測、面部偵測、面部辨識、牌照辨識、識別「遺留」物件、監測物件(例如，以防止盜竊)及商業智慧。然而，將理解，此項技術中已知之其他視訊分析功能亦可由視訊分析模組224實施。

視訊管理模組232接收影像資料且對影像資料執行與視訊傳輸、播放

及/或儲存相關之處理功能。舉例而言，視訊管理模組232可處理影像資料以准許根據頻帶要求及/或容量而傳輸影像資料。視訊管理模組232亦可根據將播放視訊之一用戶端裝置164之播放能力(諸如用戶端裝置164之處理能力及/或其顯示器之解析度)而處理影像資料。視訊管理模組232亦可根據視訊擷取與播放系統100內用於儲存影像資料之儲存容量而處理影像資料。

將理解，根據某些實例性實施例，視訊處理模組子集216可包含視訊分析模組224及視訊管理模組232中之僅一者。

操作模組集合200進一步包含一儲存模組子集240。舉例而言且如所圖解說明，儲存模組子集240包含一視訊儲存模組248及一後設資料儲存模組256。視訊儲存模組248儲存可係由視訊管理模組處理之影像資料之影像資料。後設資料儲存模組256儲存自視訊分析模組224輸出之資訊資料。

將理解，儘管將視訊儲存模組248及後設資料儲存模組256圖解說明為單獨模組，但其可實施於同一硬體儲存裝置內，藉此邏輯規則經實施以將所儲存視訊與所儲存後設資料分開。在其他實例性實施例中，視訊儲存模組248及/或後設資料儲存模組256可實施於其中可實施一分散式儲存方案之複數個硬體儲存裝置內。

該操作模組集合進一步包含可操作以接收影像資料且播放影像資料作為一視訊之至少一個視訊播放模組264。舉例而言，視訊播放模組264可實施於一用戶端裝置164上。

集合200之操作模組可實施於影像擷取裝置108、處理器具148、工作站156及用戶端裝置164中之一或多者上。在某些實例性實施例中，一操作模組可完全實施於一單個裝置上。舉例而言，視訊分析模組224可完全

實施於工作站156上。類似地，視訊管理模組232可完全實施於工作站156上。

在其他實例性實施例中，集合200之一操作模組之某些功能性可部分地實施於一第一裝置上，而一操作模組之其他功能性可實施於一第二裝置上。舉例而言，視訊分析功能性可在一影像擷取裝置108、處理器具148及工作站156中之一或多者之間劃分。類似地，視訊管理功能性可在一影像擷取裝置108、處理器具148及工作站156中之一或多者之間劃分。

現在參考圖1C，其中圖解說明根據一項特定實例性實施例之視訊擷取與播放系統100之一操作模組集合200的一方塊圖，其中視訊分析模組224、視訊管理模組232及儲存裝置240完全實施於一或多個影像擷取裝置108上。因此，視訊擷取與播放系統100可不需要一工作站156及/或一處理器具148。

將瞭解，允許影像資料處理模組子集實施於視訊擷取與播放系統100之一單個裝置上或各種裝置上實現在構建系統100時之靈活性。

舉例而言，可選擇使用具有某些功能性之一特定裝置，其中另一裝置缺乏彼等功能性。此可在整合來自不同方(例如，製造商)之裝置或改裝一現有視訊擷取與播放系統時係有用的。

現在參考圖2，其中圖解說明用於對由一視訊擷取裝置108擷取之一視訊之一或多個影像圖框執行視訊分析之一方法272之一實例性實施例的一流程圖。視訊分析可由視訊分析模組224執行以判定所擷取影像或視訊及/或存在於視訊中所擷取之場景中之視覺物件之性質或特性。

在300處，將視訊之至少一個影像圖框分段成前景區及背景區。該分段將所擷取場景中對應於移動物件(或先前移動物件)的影像圖框之區與場

景之固定區分開。

在302處，基於300之分段而偵測由影像圖框表示之場景中之一或多個前景視覺物件。舉例而言，可將任何離散連續前景區或「斑點(blob)」識別為場景中之一前景視覺物件。舉例而言，僅將大於一特定大小(例如，像素數)之連續前景區識別為場景中之一前景視覺物件。

可進一步產生與所偵測一或多個前景區相關之後設資料。後設資料可定義前景視覺物件在影像圖框內之位置。舉例而言，位置後設資料可進一步用於產生描繪所偵測前景視覺物件之輪廓之一定界框(例如，在將視訊編碼或播放視訊時)。

可將一視覺指示符添加至影像圖框以在視覺上識別所偵測一或多個前景視覺物件中之每一者。視覺指示符可係環繞影像圖框內之一或多個前景視覺物件中之每一者之一定界框。

根據各項實例性實施例，視訊分析可以偵測所擷取場景中之物件結束。

在其他實例性實施例中，視訊分析可進一步包含在304處將在302處偵測之前景視覺物件分類。舉例而言，可執行圖案辨識以將前景視覺物件分類。可將一前景視覺物件分類為諸如一人、一轎車或一動物等種類。另外或另一選擇係，可按動作(諸如視覺物件之移動及移動方向)將一視覺物件分類。亦可判定其他分類器，諸如色彩、大小、定向等。在更多特定實例中，將視覺物件分類可包含基於面部偵測而識別一人及辨識文字，諸如一牌照。可根據共同擁有之美國專利第8,934,709號中所闡述之系統及方法執行視覺分類，該專利以全文引用方式併入本文中。

視訊分析可進一步包含在306處偵測一事件是否已發生及事件之類

型。偵測事件可基於一或多個前景視覺物件之分類與一或多個預定義規則之一比較。該事件可係異常偵測中之一事件或商業智慧，諸如是否已觸發一視訊絆網，存在於一個區中之人數，是否「遺留」場景中之一物件或是否已移除場景中之一物件。

現在參考圖3A，其中圖解說明根據一項實例性實施例之視訊分析模組之一操作子模組集合400之一方塊圖。視訊分析模組400包含用於執行各種任務之若干個模組。舉例而言，視訊分析模組400包含用於偵測出現於視訊擷取裝置108之視野中之物件之一物件偵測模組404。物件偵測模組404可採用任何已知物件偵測方法，舉例而言，諸如運動偵測及斑點偵測。物件偵測模組404可包含標題為「**Methods and Systems for Detecting Objects of Interest in Spatio-Temporal Signals**」之共同擁有之美國專利第7,627,171號中所闡述之系統且使用其中所闡述之偵測方法，該專利之全部內容以引用方式併入本文中。

視訊分析模組400亦可包含連接至物件偵測模組404之一物件追蹤模組408。物件追蹤模組408可操作以在時間上使由物件偵測模組404偵測之一物件之例項相關聯。物件追蹤模組408可包含標題為「**Object Matching for Tracking, Indexing, and Search**」之共同擁有之美國專利第8,224,029號中所闡述之系統且使用其中所闡述之方法，該專利之全部內容以引用方式併入本文中。物件追蹤模組408產生對應於其追蹤之視覺物件之後設資料。後設資料可對應於視覺物件之表示物件之外觀或其他特徵之特徵符號。後設資料可傳輸至後設資料資料庫256以進行儲存。

視訊分析模組400亦可包含連接至物件追蹤模組408之一時間物件分類模組412。時間物件分類模組412可操作以藉由考量一物件隨時間之外

觀而根據物件之類型將物件分類(例如，人類、交通工具、動物)。換言之，物件追蹤模組408針對多個圖框追蹤一物件，且時間物件分類模組412基於物件在多個圖框中之外觀而判定物件之類型。舉例而言，對一人走路之方式之步態分析可適用於將一人分類，或對一人之腿之分析可適用於將一自行車運動員分類。時間物件分類模組412可組合關於一物件之軌跡之資訊(例如，軌跡是平滑的還是混亂的，物件是正在移動還是無運動)與由一物件分類模組416(下文詳細地闡述)做出之跨多個圖框求平均之對分類之信賴度。舉例而言，由物件分類模組416判定之分類信賴度值可基於物件之軌跡之平滑度而調整。時間物件分類模組412可將一物件指派為一未知種類直至視覺物件由物件分類模組分類充分次數且已收集預定數目個統計為止。在將一物件分類時，時間物件分類模組412亦可考慮到物件在視野中存在多長時間。時間物件分類模組可基於上文所闡述之資訊而做出關於一物件之種類之一最後判定。時間物件分類模組412亦可使用一滯後方法來改變一物件之種類。更具體而言，可設定用於將一物件之類別自未知轉變為一明確種類之一臨限值，且彼臨限值可大於用於相反轉變之一臨限值(例如，自一人類轉變為未知)。時間物件分類模組412可產生與一物件之種類相關之後設資料，且後設資料可儲存於後設資料資料庫256中。時間物件分類模組412可聚合由物件分類模組416做出之分類。

視訊分析模組400亦包含較佳地直接或間接連接至物件偵測模組404之物件分類模組416。與時間物件分類模組412相比而言，物件分類模組416可基於一視覺物件之一單個例項(例如，單個影像)而判定物件之類型。至物件分類模組416之輸入較佳地係一影像圖框之所關注視覺物件位於其中之一子區域而非整個影像圖框。將影像圖框之一子區域輸入至物件

分類模組416之一益處係不必分析全部場景以進行分類，藉此需要較少處理能力。亦可包含用以捕捉明顯分類之其他初級模組(諸如一基於啟發法之模組)以進一步簡化物件分類模組416之複雜性。

在一替代配製中，物件分類模組416放置於物件偵測模組404之後及物件追蹤模組408之前，使得物件分類在物件追蹤之前發生。在另一替代配製中，物件偵測模組404、物件追蹤模組408、物件時間分類模組412及物件分類模組416係互相關的，如上文參考中所闡述。

物件分類模組416包含若干個物件分類器，如圖3B之方塊圖中所繪示。舉例而言，物件分類模組416可包含判定一所偵測物件之一影像是否對應於一全人體之一全人體分類器424、判定一所偵測物件之一影像是否對應於一人體軀幹之一人體軀幹分類器428及判定一所偵測物件之一影像是否對應於一交通工具之一交通工具分類器432。物件分類模組416可包含任何數目個不同分類器，且如下文更詳細地闡述，一使用者可甚至在攝影機系統經部署且起作用時為物件分類模組416建立物件之新種類。換言之，物件分類模組416係現場可訓練的。

物件分類器可操作以基於一物件之特徵(例如，外觀特性)而將物件分類。舉例而言，全人體分類器424接收對應於一物件之特徵之資料(亦即，一輸入圖案 X)且判定物件是否對應於一全人體。在物件分類模組416將一物件分類之後，表示物件之種類及物件之特徵之後設資料可儲存於後設資料資料庫256中。

現在將更詳細地闡述可由物件分類模組416使用之特徵。下文所闡述之一訓練演算法自一特徵集合 $F=\{f_1, f_2, \dots, f_n\}$ 選擇一特徵子集 $\hat{F}=\{f_{k1}, f_{k2}, \dots, f_{km}\}$ 。輸入圖案 X 係由 $\hat{F}$ 之元素構成。 $\hat{F}$ 之元素可被視為一物

件之一影像區域 R 之某一變換。因此， X 可採取以下形式：

$$X = \begin{pmatrix} \hat{f}_1 = f_{k1}(R) \\ \hat{f}_2 = f_{k2}(R) \\ \vdots \\ \hat{f}_m = f_{km}(R) \end{pmatrix}$$

一物件之特徵 $\hat{f}_1, \hat{f}_2, \dots, \hat{f}_m$ 可對應於任何數目個外觀特性，諸如但不限於縱橫比、色彩、邊緣定向及經正規化飽和。此外，特徵 $\hat{f}_1, \hat{f}_2, \dots, \hat{f}_m$ 可表示外觀特性之特徵向量(例如，其中直方圖組格對應於向量分量之直方圖)且可由一或多個物件分類器用於判定物件之種類(例如，類型)。舉例而言，可針對一物件之影像之不同區域(例如，子視窗)建構物件之邊緣定向之直方圖。換言之，可將一物件之一影像劃分成若干子視窗，且可針對該等子視窗之每一像素計算邊緣定向。可使用一可導引濾波器(例如，沿多個方向使用一高斯導數濾波器)衍生出一像素之邊緣定向。使用一可導引濾波器允許將主導方向指派給一子視窗之像素，且允許針對子視窗建構方向之一直方圖。舉例而言，針對一給定像素，一可導引濾波器可沿多個方向用於產生多個回應，且將對應於最大方向衍生回應之方向指派為像素之方向。

物件分類器中之一者之分類問題可通常由一分類器函數 $\Gamma(X)$ 定義，其中當 $\Gamma(X) > 0$ 時將由輸入圖案 X 表示之一視覺物件宣告為物件種類之一成員，或當 $\Gamma(X) < 0$ 時將其宣告為物件種類之一非成員。通常，分類器函數 $\Gamma(X)$ 係以一參數集合參數化，且輸入圖案 X 係由上文所闡述之特徵組成。針對每一所關注物件種類訓練一特定分類器 $\Gamma_c(X)$ 。由圖3A之物件分類模組416表示之多種類分類模型可在數學上定義為如下：

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_c\}$$

$$\omega = \omega_c: (\Gamma_c(X) > 0 \text{ 且 } \Gamma_c(X) > \Gamma_u(X) \forall u \in \{1, 2, \dots, C\}, u \neq c)$$

其中 ω 表示一物件種類，且 Ω 表示所有物件種類之集合。

可藉由定義規則(例如，視覺物件之大小及縱橫比)而構建用於一給定視覺物件種類之一分類器函數 $R(X)$ 。可藉由使用訓練資料應用機器學習而進一步訓練分類器函數。如此項技術中已知，訓練一分類器尋求進一步細化彼分類器之規則，使得其可更準確地將一給定視覺物件分類。訓練資料可包含正向訓練實例及/或負向訓練實例。一正向訓練實例係指已確認為屬於一特定物件種類之一視覺物件之一例項。正向訓練實例用於訓練一分類器以細化其規則以便更準確地將一給定視覺物件正向地分類為屬於彼正向訓練實例之種類內。一負向訓練實例係指不屬於一特定物件種類之一視覺物件或其他視覺表示之一例項。負向訓練實例可係被一分類器錯誤分類為屬於一特定物件種類之一視覺物件之一實例。負向訓練實例用於訓練一分類器。

用於訓練物件分類器之機器學習可係此項技術中已知之任何適當機器學習技術，諸如但不限於迴旋神經網路、支援向量機器、決策樹、隨機森林、級聯分類器。

可監督對一物件分類器之訓練。在受監督訓練中，正向訓練實例及/或負向訓練實例由一人類使用者確認。舉例而言，在一大批次影像當中，一或多個人類使用者個別地檢查每一影像且將每一影像標記為表示屬於一種類(例如，人、交通工具、動物)之一視覺物件或標記為不含一視覺物件。

亦可不監督對一物件分類器之訓練。在不受監督訓練中，使用一基

礎分類器來最初將一或多個視覺物件(諸如由物件偵測模組404偵測之物件)分類。視覺物件及由基礎分類器判定之分類結果(例如，視覺物件屬於一特定物件種類之一正向判定)可用作用於進一步訓練基礎分類器之一正向訓練實例。其中未偵測到物件之影像資料亦可用作用於訓練物件分類器之負向訓練實例。在不受監督訓練中，用作正向訓練實例或用作負向訓練實例之影像資料並不由一人類使用者檢查。

本文中一基礎分類器係指已透過定義規則組態及/或透過應用機器學習訓練以執行一特定程度之物件分類但可透過使用電腦實施之視覺機器語言之進一步訓練而進一步最佳化之一物件分類器。

現在參考圖4，其中圖解說明用於進一步訓練一基礎分類器之一方法500之一流程圖。將理解，儘管針對訓練一單個基礎分類器而圖解說明方法500，但方法500可應用於訓練複數個並聯基礎分類器。舉例而言且如本文中其他處所闡述，一物件分類模組416可包含複數個物件分類器，每一分類器可操作以判定一視覺物件是否屬於一特定種類類型。因此，物件分類模組416之複數個物件分類器可基於提供至該物件分類模組之訓練實例而一起訓練。舉例而言，係一特定種類之一前景視覺物件之一訓練實例可用作一分類器之屬於同一種類之一正向訓練實例。

在504處，提供一基礎分類器。基礎分類器可係可使用視覺訓練實例透過應用機器學習而進一步訓練之任何物件分類器。

在508處，可接收一或多個訓練實例。該等訓練實例可係自動準備或在受監督條件下準備之正向訓練實例及/或負向訓練實例。

在512處，藉由使用在508處接收之訓練實例作為輸入將機器學習應用於基礎分類器而進一步訓練基礎分類器。

將理解，在某些實施例中，重複步驟508及512使得基礎分類器之更新遵循一反覆程序。亦即，一第一批次複數個訓練實例可在第一反覆中應用於藉由機器學習而訓練基礎分類器。一第二批次複數個訓練實例可在後續第二反覆中進一步應用於藉由機器學習而進一步訓練分類器。

在516處，將在步驟508及512之後訓練之基礎分類器部署於視野中以用於將前景視覺物件分類。

在某些實例中，依據步驟508及512對基礎分類器之訓練可在於516處部署經訓練分類器之前執行。

在其他實例中，在步驟508及512處對一基礎分類器之訓練可在物件分類器已部署於視野中時執行。訓練實例可係在一視訊擷取裝置部署於視野中時存在於彼裝置之視野中之現實世界物件之視覺表示。舉例而言，基礎分類器可經最初部署且在部署期間從自視野偵測之前景視覺物件508逐步訓練。

可以一受監督方式(例如，由一人類使用者在視覺上檢查)或以一不受監督方式(例如，由一電腦實施之物件分類器分類)將用作訓練實例之視覺物件識別為屬於一種類。

現在參考圖5，其中圖解說明根據一項實例性實施例之用於進一步訓練一基礎分類器之一經改良電腦實施之方法540之一流程圖。將理解，儘管針對訓練一單個基礎分類器而圖解說明方法540，但方法540亦可應用於訓練複數個並聯基礎分類器。舉例而言且如本文中其他處所闡述，一物件分類模組416可包含複數個物件分類器，每一分類器可操作以判定一視覺物件是否屬於一特定種類。因此，物件分類模組416之複數個物件分類器可基於提供至該物件分類模組之訓練實例而一起訓練。舉例而言，係一

特定種類之一視覺物件之一訓練實例可用作一分類器之屬於同一種類之一正向訓練實例。

在504處，提供一基礎分類器。基礎分類器可係可使用視覺物件訓練實例透過應用機器學習而進一步最佳化之任何物件分類器。

在544處，在表示一場景之影像資料內偵測一前景視覺物件。本文中一場景係指跨一時間間隔在一視訊擷取裝置之視野內擷取之視覺表示。視訊擷取裝置跨此時間間隔係固定的，使得其視野保持不變。因此，跨該時間間隔擷取之場景亦保持不變，但場景內之物件(例如人類、交通工具、其他物件)可跨彼時間間隔改變。場景之視覺表示可係由視訊擷取裝置跨該時間間隔產生之影像資料之影像圖框。

前景視覺物件亦可由一人類操作者或由一電腦實施之模組正向地分類為屬於一特定種類。所偵測之前景視覺物件位於場景之一子區域內。舉例而言，場景之子區域可對應於所偵測前景視覺物件位於其中的影像資料之一影像圖框之一部分。舉例而言，場景之子區域可對應於藉由由物件偵測模組404繪製以用於在視覺上識別所偵測前景視覺物件之定界框劃定界線的影像圖框之子區域。

在548處，判定所偵測視覺物件之一背景模型。一背景模型係場景之一視覺表示或其之一子區域但其中任何前景視覺物件皆不存在於該場景或該子區域中。所偵測前景視覺物件之背景模型係所偵測之前景視覺物件位於其中的場景之子區域之背景模型。

舉例而言，在於544處偵測之前景視覺物件係一人類且場景之子區域對應於彼人類位於其中的一空間之一區之情況下，彼子區域之背景模型表示不存在彼人類或任何其他類型的該房間之彼區。

舉例而言，在於544處偵測之前景視覺物件係一交通工具且場景之子區域對應於彼交通工具位於其中的一停車場之一部分之情況下，彼子區域之背景模型表示不存在彼交通工具或任何其他交通工具的該停車場之彼部分。

在552處，視情況藉由使用在544處偵測之前景視覺物件作為一正向訓練實例來將機器學習應用於基礎分類器而進一步訓練基礎分類器。

在556處，藉由使用所偵測前景視覺物件之背景模型作為一負向訓練實例來將機器學習應用於基礎分類器而進一步訓練基礎分類器。

可針對所偵測及/或分類之複數個視覺物件重複步驟544至556。針對在544處偵測之每一視覺物件，在548處判定彼視覺物件位於其中的一場景之子區域所特有之一背景模型且在556處將其應用於訓練基礎分類器。

在其他實例中，可藉由使用一批次複數個訓練實例來將機器學習應用於基礎分類器而訓練基礎分類器。此批次包含在複數個不同場景之子區域中偵測之前景視覺物件之複數個背景模型。

將理解，在某些實施例中，重複步驟544及556使得基礎分類器之更新遵循一反覆程序。亦即，亦即，一第一批次一或多個訓練實例可在第一反覆中應用於藉由機器學習而訓練基礎分類器。一第二批次複數個訓練實例可在一後續第二反覆中進一步應用於藉由機器學習而進一步訓練在第一反覆之後經訓練之基礎分類器。

在516處，將在步驟556及視情況步驟552之後訓練之基礎分類器部署於視野中以用於將額外前景視覺物件分類。

如本文中其他處所闡述，對基礎分類器之訓練可在部署經訓練分類器之前或在物件分類器已部署於視野中時執行。

圖6A至圖6F展示在場景之子區域中偵測之前景視覺物件及其對應背景模型。舉例而言，圖6A展示行走於一人行道之一路段上之一人。行走之人係所偵測之前景視覺物件。圖6B展示圖6A之視覺物件之背景模型。將瞭解，該背景模型展示不存在行走之人或任何其他前景視覺物件的人行道之同一路段。

圖6C展示沿一段樓梯下行之一人。該人係所偵測之前景視覺物件。圖6D展示圖6C之前景視覺物件之背景模型。將瞭解，該背景模型展示不存在人或任何其他前景視覺物件之同一段樓梯。

圖6E展示在一道路路段上駕駛之一交通工具。該交通工具係所偵測之前景視覺物件。圖6F展示圖6E之前景視覺物件之背景模型。將瞭解，該背景模型展示不存在交通工具或任何其他前景視覺物件之同一道路路段。

根據各項實例性實施例，自一歷史影像圖框判定一所偵測視覺物件之一背景模型。在形成由視訊擷取裝置擷取之視訊之影像資料之一影像圖框序列之一當前影像圖框的一給定子區域內偵測一前景視覺物件。一歷史影像圖框係該影像圖框序列中之一先前影像圖框，其中該前景視覺物件及任何其他前景視覺物件不存在於彼先前影像圖框中。在此情形中，當前影像圖框及歷史影像圖框表示同一場景。亦即，在歷史影像圖框與當前影像圖框之時間之間視訊擷取裝置係固定的(亦即，不移動)，以使視訊擷取裝置擷取同一場景。自歷史影像圖框裁剪對應於前景視覺物件位於其中的當前影像圖框之子區域的歷史影像圖框之一給定子區域。以此方式裁剪之歷史影像圖框係所偵測前景視覺物件之背景模型。在556處提供此經裁剪歷史影像圖框作為用於進一步訓練基礎分類器之一負向訓練實例。

根據各項實例性實施例，可最初建構整個場景之一完整背景模型。然後，可自完整背景模型提取場景之一給定子區域之背景模型。

舉例而言，在一較不忙碌場景(諸如其中存在前景視覺物件之一低發生率之一個場景)中，可將完全不含前景物件之一單個歷史影像圖框用作完整背景模型。

在一較忙碌場景中，在場景內在任何時間可總是存在至少一個前景視覺物件。針對此類場景，可藉由聚合來自複數個歷史影像圖框之不同子區域以形成完整背景模型而建構完整背景模型。

根據一項實例，選擇複數個歷史影像圖框。此等歷史影像圖框中之每一者含有不含任何前景物件的影像圖框之至少一個子區域。

判定每一選定歷史影像圖框之不含任何前景物件之一或多個子區域之座標。可自此等子區域之各別歷史影像圖框裁剪此等子區域。

然後，聚合自複數個歷史影像裁剪之子區域以形成一經聚合影像。可藉由適當地選擇複數個歷史影像圖框使得不含任何前景物件的此等圖框之子區域整體地涵蓋整個場景而獲得表示整個場景之一經聚合影像。因此，經聚合影像形成場景之一完整背景模型。舉例而言，可根據此項技術中已知之拼接方法將自複數個歷史影像裁剪之影像子區域拼接在一起以形成經聚合影像。

因此，在偵測到一場景之一給定子區域內之一前景視覺物件之後，可藉由裁剪對應於其中偵測到該視覺物件之給定子區域的經聚合影像之一子區域而獲得彼子區域之背景模型。

圖7A展示表示係一廣場之一實例性場景之一第一全歷史影像圖框。將瞭解，涵蓋餐飲區及草地區之部分之第一子區域700不含任何前景視覺

物件。因此，第一子區域700可用作待聚合以形成完整背景模型之子區域中之一者。然而，涵蓋階梯之第二子區域708有一人位於其中。由於第一全歷史影像圖框中之此第二子區域708包含一前景視覺物件，因此其無法用於構建完整背景模型。

圖7B展示表示廣場之同一場景之一第二全歷史影像圖框。第二全歷史影像圖框係在比第一全歷史影像稍後之一時間點處擷取。將瞭解，第二全歷史影像圖框中之第二子區域708不含一前景視覺物件。第一全歷史影像圖框中在階梯上之人現在已完全自階梯下來。因此，第二全歷史影像圖框中之此第二子區域708可用作待聚合以形成完整背景模型之子區域中之一者。可以相同方式判定適用於形成完整背景模型的場景之其他子區域。

現在參考圖8，其中圖解說明根據一替代實例性實施例之用於進一步訓練一基礎分類器之一經改良電腦實施之方法558之一流程圖。替代實例性方法558包含與方法540相同之步驟，但亦包含額外步驟560及564。

在560處，提供一場景之一經錯誤分類子區域。一場景之一經錯誤分類子區域係指其中一物件分類器在一子區域實際上不含一特定種類之任何物件時錯誤地將該子區域分類為含有屬於彼種類之一物件的該子區域。

可在其中由一物件分類器分類之物件由識別由物件分類器做出之任何錯誤分類之一人類檢查之一受監督環境中判定一經錯誤分類子區域。

可在一部分受監督或完全不受監督環境中判定一經錯誤分類子區域。在一項實例中，可將其中不存在物件的影像圖框之子區域饋送至一物件分類器。由物件分類器進行之子區域包含屬於一特定種類之一物件之任何分類(除分類為背景外)將係錯誤的且被識別為一經錯誤分類子區域。

其中識別一經錯誤分類子區域之場景可係與其中在544處偵測一前景

視覺物件之場景相同之場景。另一選擇係，經錯誤分類子區域之場景可不同於其中偵測前景視覺物件之場景。

在564處，藉由使用經錯誤分類子區域作為一負向訓練實例來將機器學習應用於基礎分類器而進一步訓練基礎分類器。

在516處，部署自所偵測視覺物件之背景模型、經錯誤分類子區域及視情況所偵測視覺物件訓練之分類器以用於將其他所偵測視覺物件分類。

現在參考圖9，其中圖解說明根據一項實例性實施例之用於一基礎分類器之場景特定訓練之一經改良電腦實施之方法600之一流程圖。將理解，實例性方法600之眾多步驟與實例性方法540之步驟類似或相同，且關於實例性方法540所提供之說明亦可應用於實例性方法600。將理解，場景特定方法600亦可根據替代實例性方法560而應用。

在504處，提供一基礎分類器。

在提供基礎分類器之後，對基礎分類器之訓練開始。專門針對一當前現實世界場景訓練基礎分類器。當前場景可對應於定位於一特定位置處且沿一特定方向定向之一特定攝影機之視野。

在544處，在表示當前場景之影像資料內偵測一前景視覺物件。

在548處，判定所偵測物件之一背景模型。

在552處，藉由使用在544處自當前場景偵測之前景視覺物件作為一正向訓練實例來將機器學習應用於基礎分類器而視情況訓練基礎分類器。

在556處，藉由使用在548處判定的前景視覺物件之背景模型作為一負向訓練實例來將機器學習應用於基礎分類器而訓練基礎分類器。

在516處，部署基於當前場景之前景視覺物件及/或背景模型而訓練之基礎分類器以用於將存在於當前場景中之物件分類。

將理解，只要當前場景保持不變，便可重複步驟544至556以便藉由使用存在於當前場景中之複數個訓練實例來應用機器學習而進一步訓練基礎分類器。如本文中其他處所闡述，可重複步驟544至556使得基礎分類器之更新遵循一反覆程序。

在608處，判定當前場景是否已改變。當前場景之此一改變可由於擷取場景之攝影機之位置之一改變而發生。此一改變亦可由於擷取場景之攝影機之定向之一改變而發生。此一改變亦可進一步由於擷取場景之攝影機之一設定之一改變(諸如由攝影機應用之變焦或攝影機之一操作模式之一顯著改變(例如自正常光模式切換為低光模式))而發生。

若在608處場景保持不變，則方法600可返回至544以對場景內之額外視覺物件進行偵測及分類。另一選擇係，方法600可返回至516以繼續部署針對當前場景依據步驟544至556訓練之物件分類器。

若在608處場景改變，則方法進行至步驟616以至少部分地恢復為基礎分類器。在某些實例中，當存在場景之一改變時，將當前在516處部署之物件分類器完全恢復回基礎分類器。

在於步驟616處恢復回基礎分類器之後，可將自場景之改變產生之新場景設定為當前場景。然後，方法600可返回至544以對存在於「新」當前場景中之前景視覺物件進行偵測及分類。此等物件及/或對應於彼等物件之背景模型可應用於在步驟616之恢復之後更新基礎分類器。

將瞭解，恢復為基礎分類器可在其中一初始場景與一後續場景之特性顯著不同使得根據初始場景之特性對基礎分類器之訓練不適用於後續場景之情況中係有用的。恢復回基礎分類器允許專門針對後續場景之特性而訓練分類器。

實驗

根據一項實驗，在使用不同訓練實例集合訓練時評估一基礎分類器之效能 (Alex Krizhevsky 、 Ilya Sutskever 、 Geoffrey Hinton 之「ImageNet Classification with deep convolution neural networks」(NIPS 2012)中所闡述之稱為「AlexNet」之深度迴旋神經網路之一特定架構)。

自VIRAT資料集(<http://www.viratdata.org>)獲得訓練實例。此資料集包含來自各種固定式攝影機之超過300個視訊。自VIRAT資料集提取人類種類之視覺物件及交通工具種類之視覺物件且將其用作一第一訓練實例集合。針對用作一訓練實例之每一視覺物件判定一背景模型。此等背景模型用作一第二訓練實例集合。

除人類種類及交通工具種類外，亦自VIRAT資料集提取屬於背景種類之訓練實例。為了產生背景種類訓練實例，準備不含人類種類之一前景視覺物件或交通工具種類之一前景視覺物件之影像樣本。每一影像樣本係存在於VIRAT資料集中之視訊之一影像圖框之一經裁剪部分。使用一簡單物件分類器(諸如不基於神經網路分類器之一個分類器)來將此等影像樣本分類。在簡單分類器將影像樣本中之任一者分類為含有落在人類種類或交通工具種類內之一視覺物件時發生一錯誤分類。一第三訓練實例集合中包含此等經錯誤分類影像樣本。

AlexNet分類器經提供作為待藉由自VIRAT資料集提取之訓練實例訓練之一基礎分類器。使用來自伯克利視覺與學習中心(Berkeley Vision and Learning Center)之Caffe深度學習框架(Caffe deep learning framework)(可在caffe.berkeleyvision.org處獲得)將正向及負向訓練實例

應用於訓練基礎分類器。在一Tesla K80 GPU上執行對基礎分類器之更新。

在實驗之一第一部分中，藉由應用第一實例集合作為正向訓練實例(100個正向訓練實例)及藉由應用第二訓練實例集合作為負向訓練實例(100個負向訓練實例)而訓練基礎分類器。對基礎分類器之此訓練產生一第一經訓練測試分類器。

在實驗之一第二部分中，藉由應用第一實例集合作為正向訓練實例(100個正向訓練實例)及藉由應用第三訓練實例集合作為負向訓練實例(100個負向訓練實例)而訓練基礎分類器。對基礎分類器之此訓練產生一第二經訓練測試分類器。

在實驗之一第三部分中，藉由應用第一實例集合作為正向訓練實例(100個正向訓練實例)及第二訓練實例集合與第三訓練實例集合之一混合作為負向訓練實例而訓練基礎分類器。更精確地，應用來自第二集合之50個訓練實例及來自第三集合之50個訓練實例作為用於訓練基礎分類器之負向訓練實例。對基礎分類器之此訓練產生一第三經訓練測試分類器。

部署第一經訓練測試分類器、第二經訓練測試分類器及第三經訓練測試分類器中之每一者以用於對來自一內部視訊資料集之一測試視訊集合之物件分類。度量在部署該等分類器中之每一者時之錯誤率。在將一視覺物件錯誤分類時或在將一背景影像(例如，不存在前景視覺物件)分類為在人類種類或交通工具種類中之一視覺物件時考量出現一錯誤。

表1係展示第一經訓練測試分類器在經部署以用於將測試視訊集合中所含之前景視覺物件分類時之效能之一混淆矩陣。

表1：

		經分類(%)		
		人類	交通工具	背景
實際	人類	36.26	9.58	3.26
	交通工具	0.94	26.78	0.29
	背景	0.14	0.15	22.60

表2係展示第二經訓練測試分類器在經部署以用於將測試視訊集合中所含之前景視覺物件分類時之效能之一混淆矩陣。

表2：

		經分類(%)		
		人類	交通工具	背景
實際	人類	38.80	8.56	1.74
	交通工具	4.49	23.36	0.16
	背景	0.32	0.14	22.42

表3係展示第三經訓練測試分類器在經部署以用於將測試視訊集合中所含之前景視覺物件分類時之效能之一混淆矩陣。

表3：

		經分類(%)		
		人類	交通工具	背景
實際	人類	43.26	4.72	1.11
	交通工具	3.49	24.39	0.14
	背景	0.34	0.12	22.43

第一經訓練測試分類器之錯誤率係14.36%，第二經訓練測試分類器之錯誤率係15.42%，且第三經訓練測試分類器之錯誤率係9.92%。

將瞭解，使用前景視覺物件之背景模型來訓練基礎分類器(第一經訓練測試分類器及第三經訓練測試分類器)展現比其中不使用背景模型作為

訓練實例之第二經訓練測試分類器低之錯誤率。較低錯誤率係經改良效能之一指示符。更顯著地，將瞭解，一起使用前景視覺物件之背景模型及背景種類物件之一組合作為負向訓練實例展現顯著改良效能(對比第二經訓練測試分類器之35.6%較低錯誤率)。

在不受一特定理論束縛之情況下，使用所偵測視覺物件之背景模型作為負向訓練實例來訓練一分類器可減小分類器將一場景之原本形成場景之背景之部分之物件錯誤分類之發生率。

往回參考圖6C，將瞭解，影像中所展示之場景之子區域包含人及一路燈柱。人係一前景視覺物件且路燈柱形成場景之背景之部分。然而，當將此子區域用作一正向訓練實例時，可致使基礎分類器經訓練以將路燈柱辨識為人種類之一前景視覺物件。舉例而言，若場景之此子區域對應於將通常具有一所關注物件之一現實位置(例如，一頻繁使用之過道、路徑或道路)，則路燈柱可出現於各自用作正向訓練實例之多個子區域中。此可增加分類器將經訓練以將路燈柱辨識為人種類之一物件之一例項之概率。將子區域之背景模型用作一負向訓練實例可藉由以路燈柱形成場景之背景之部分訓練分類器而至少部分地抵消此效應。

類似地，藉由使用圖6F中所展示之背景模型訓練一分類器，該分類器經訓練以將垂直桿辨識為形成背景之部分，藉此減小將垂直桿或與其類似之物件分類為屬於一人類種類或交通工具種類之可能性。

更一般而言且在不受一特定理論束縛之情況下，使用一背景模型訓練一分類器導致一分類器經訓練以將形成一場景之背景之部分之現實物件正確地辨識為背景物件。舉例而言，在其中將經常偵測到一前景視覺物件的一場景之一子區域中，使用一背景模型、使用彼子區域之背景模型作為

一負向訓練實例可減小分類器經訓練以將形成背景之部分之物件錯誤地分類為屬於一特定種類之前景視覺物件之概率。

儘管以上說明提供實施例之實例，但將瞭解，所闡述實施例之某些特徵及/或功能易於在不背離所闡述實施例之操作之精神及原理之情況下進行修改。因此，上文所闡述之內容意欲係說明而非限制性的，且熟習此項技術者將理解可在不背離如本發明隨附之申請專利範圍中所界定之本發明之範疇之情況下做出其他變化及修改。

【符號說明】

100	視訊擷取與播放系統/影像擷取與播放系統/系統
108	視訊擷取裝置/影像擷取裝置
116	影像感測器/感測器
124	處理器
132	記憶體裝置
140	網路
148	處理器具
156	工作站
164	用戶端裝置
200	操作模組集合/集合
208	視訊擷取模組
216	影像資料處理模組子集/視訊處理模組子集
224	視訊分析模組
232	視訊管理模組
240	儲存模組子集/儲存裝置

248	視訊儲存模組
256	後設資料儲存模組/後設資料資料庫
264	視訊播放模組
400	操作子模組集合/視訊分析模組
404	物件偵測模組
408	物件追蹤模組
412	時間物件分類模組
416	物件分類模組
424	全人體分類器
428	人體軀幹分類器
432	交通工具分類器
700	第一子區域
708	第二子區域

【發明申請專利範圍】

【第1項】

一種用於訓練一電腦實施之物件分類器(object classifier)之方法，該方法包括：

偵測一視訊擷取裝置之一視野內之一場景之一子區域內之一前景(foreground)視覺物件；

判定該場景之該子區域之一背景模型，該背景模型表示其中不存在任何前景視覺物件時之該子區域；

藉由使用該子區域之該背景模型作為一第一負向訓練實例(negative training example)進行電腦實施之機器學習來訓練該物件分類器；

其中專門針對一當前場景而訓練該物件分類器，及其中在該當前場景被改變為一新場景後：

恢復(reverting)為未進行該當前場景所特有之該訓練之該物件分類器；且

藉由使用來自該新場景之背景模型進行機器學習來訓練該物件分類器。

【第2項】

如請求項1之方法，其進一步包括藉由使用該所偵測前景視覺物件作為一正向訓練實例進行機器學習來訓練該物件分類器。

【第3項】

如請求項1或2之方法，其中判定該場景之該子區域之該背景模型包括：

選擇在對應於該場景之該子區域的一歷史影像圖框之一子區域中不存在任何前景物件時擷取之該歷史影像圖框；及

自該歷史影像圖框裁剪對應於該場景之該子區域之該子區域，該經裁剪影像圖框係該場景之該子區域之該背景模型。

【第4項】

如請求項1或2之方法，其中判定該場景之該子區域之該背景模型包括：

在複數個歷史影像圖框中之每一者內判定不含任何前景物件之一或多個子區域；

聚合來自該複數個歷史影像圖框之該一或多個子區域以形成表示該整個場景之一完整背景影像；及

自該完整背景影像裁剪對應於該場景之該子區域之一子區域，該經裁剪完整背景影像係該場景之該子區域之該背景模型。

【第5項】

如請求項4之方法，其中聚合來自該複數個歷史影像圖框之該一或多個子區域包括：拼接該一或多個子區域以形成表示該整體場景之一影像。

【第6項】

如請求項1或2之方法，其中部分地使用受監督學習來準備該物件分類器。

【第7項】

如請求項1或2之方法，其中該電腦實施之機器學習係一迴旋神經網路(convolution neural network)。

【第8項】

如請求項1之方法，進一步包括藉由使用一場景之一經錯誤分類子區域(misclassified sub-region)作為一負向訓練實例之電腦實施之機器學習來訓練該物件分類器。

【第9項】

一種依據如請求項1至8中任一項之方法所訓練之電腦實施之物件分類器。

【第10項】

一種用於訓練一電腦實施之物件分類器之系統，該系統包括：

一處理器；

一電腦可讀儲存裝置，其儲存在由該處理器執行時致使該系統執行包括以下各項之操作之程式指令：

偵測一視訊擷取裝置之一視野內之一場景之一子區域內之一前景視覺物件；

判定該場景之該子區域之一背景模型，該背景模型表示其中不存在任何前景視覺物件時之該子區域；

藉由使用該子區域之該背景模型作為一第一負向訓練實例進行電腦實施之機器學習來訓練該物件分類器，其中專門針對一當前場景而訓練該物件分類器；

在該當前場景被改變為一新場景後，恢復為未進行該當前場景所特有之該訓練之該物件分類器；及

藉由使用來自該新場景之背景模型進行機器學習來訓練該物件分類器。

【第11項】

如請求項10之系統，其中該等操作進一步包括藉由使用該所偵測前

景視覺物件作為一正向訓練實例進行機器學習來訓練該物件分類器。

【第12項】

如請求項10或11之系統，其中判定該場景之該子區域之該背景模型包括：

選擇在對應於該場景之該子區域的一歷史影像圖框之一子區域中不存在任何前景物件時擷取之該歷史影像圖框；

自該歷史影像圖框裁剪對應於該場景之該子區域之該子區域，該經裁剪影像圖框係該場景之該子區域之該背景模型。

【第13項】

如請求項10或11之系統，其中判定該場景之該子區域之該背景模型包括：

在複數個歷史影像圖框中之每一者內判定不含任何前景物件之一或多個子區域；

聚合來自該複數個歷史影像圖框之該一或多個子區域以形成表示該整個場景之一完整背景影像；及

自該完整背景影像裁剪對應於該場景之該子區域之一子區域，該經裁剪完整背景影像係該場景之該子區域之該背景模型。

【第14項】

如請求項13之系統，其中聚合來自該複數個歷史影像圖框之該一或多個子區域包括：拼接該一或多個子區域以形成表示該整體場景之一影像。

【第15項】

如請求項10或11之系統，其中部分地使用受監督學習來準備該物件

分類器。

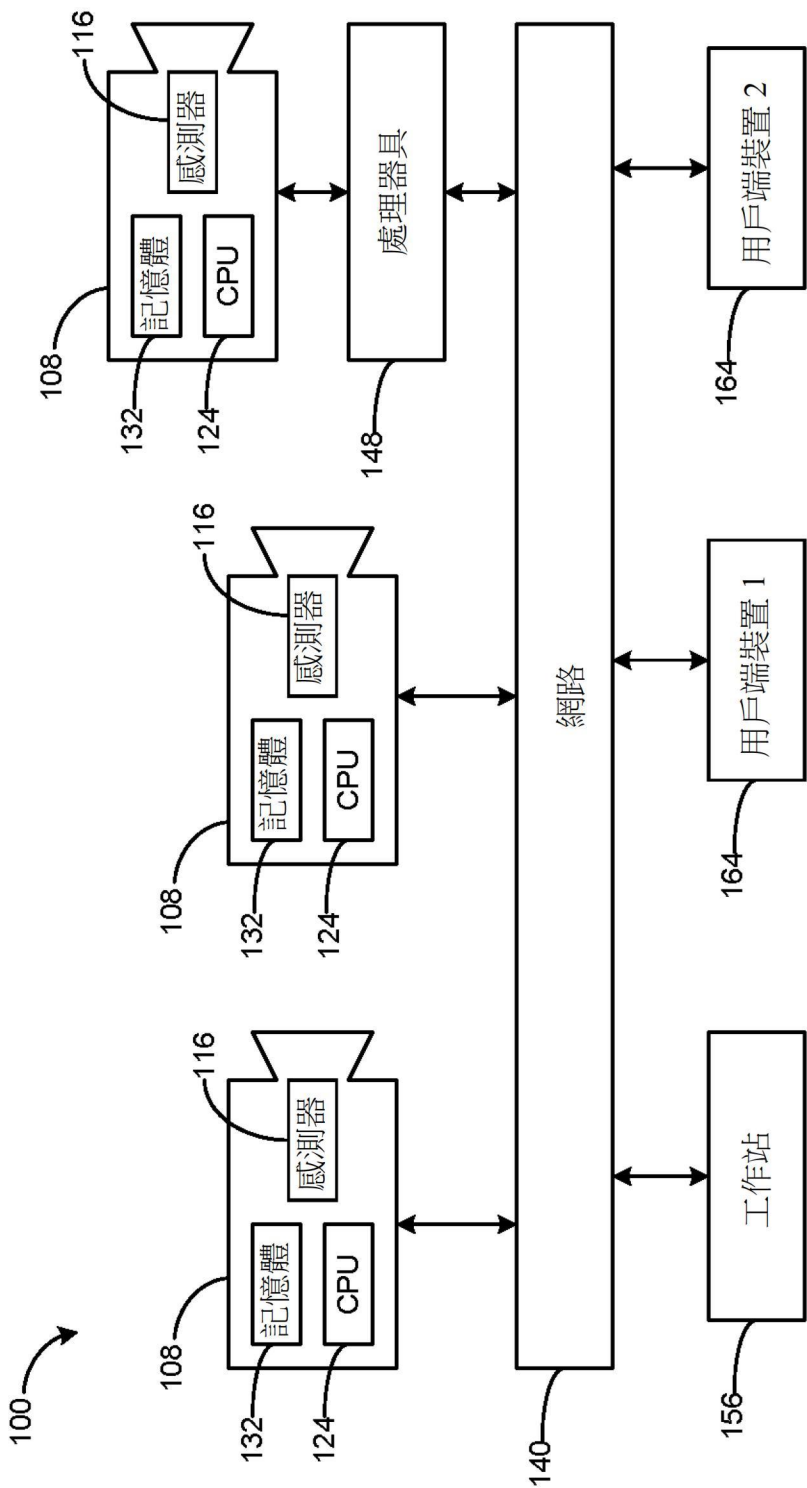
【第16項】

如請求項10或11之系統，其中該電腦實施之機器學習係一迴旋神經網路。

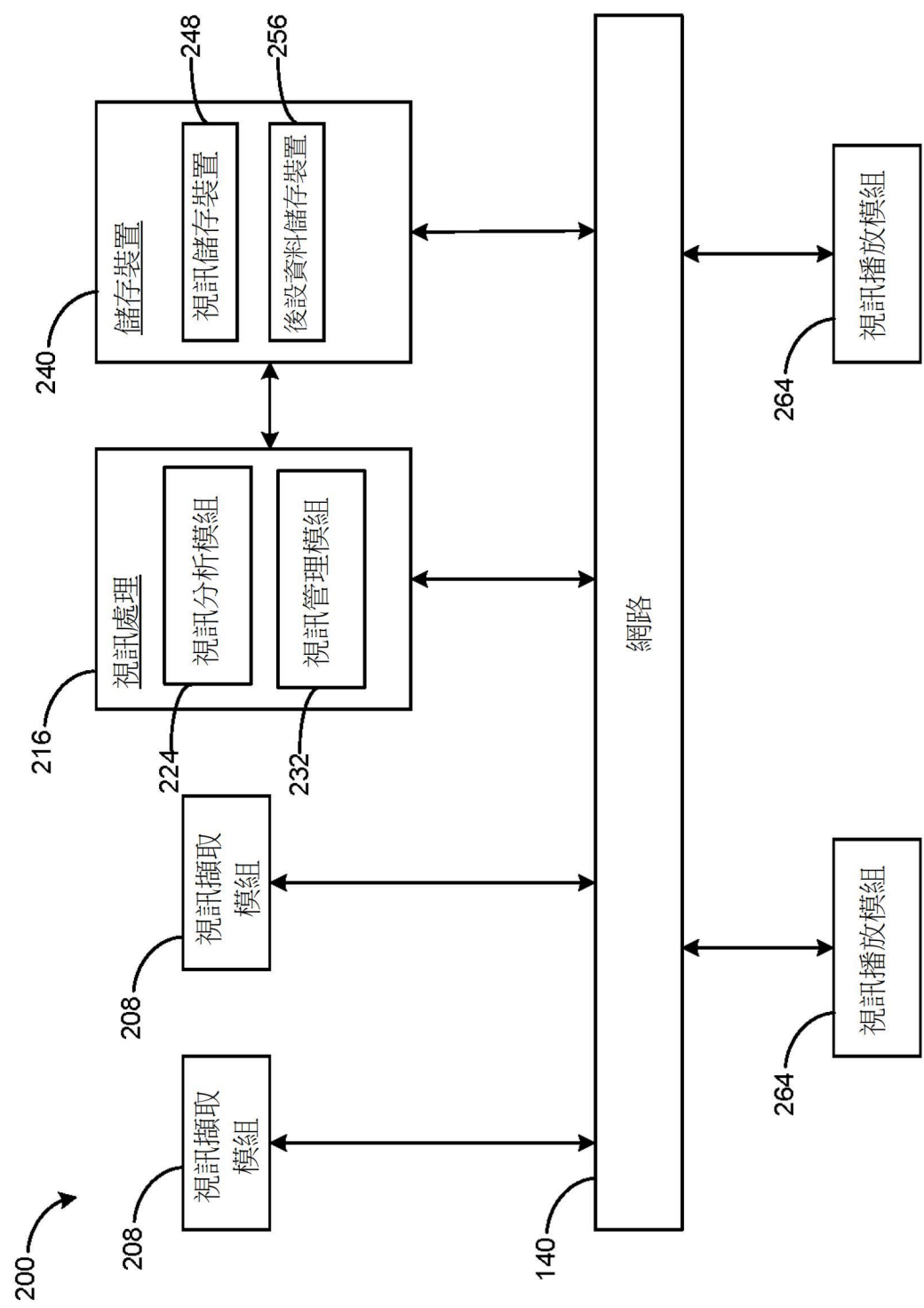
【第17項】

如請求項10之系統，其中該等操作進一步包括藉由使用一場景之一經錯誤分類子區域作為一負向訓練實例之電腦實施之機器學習來訓練該物件分類器。

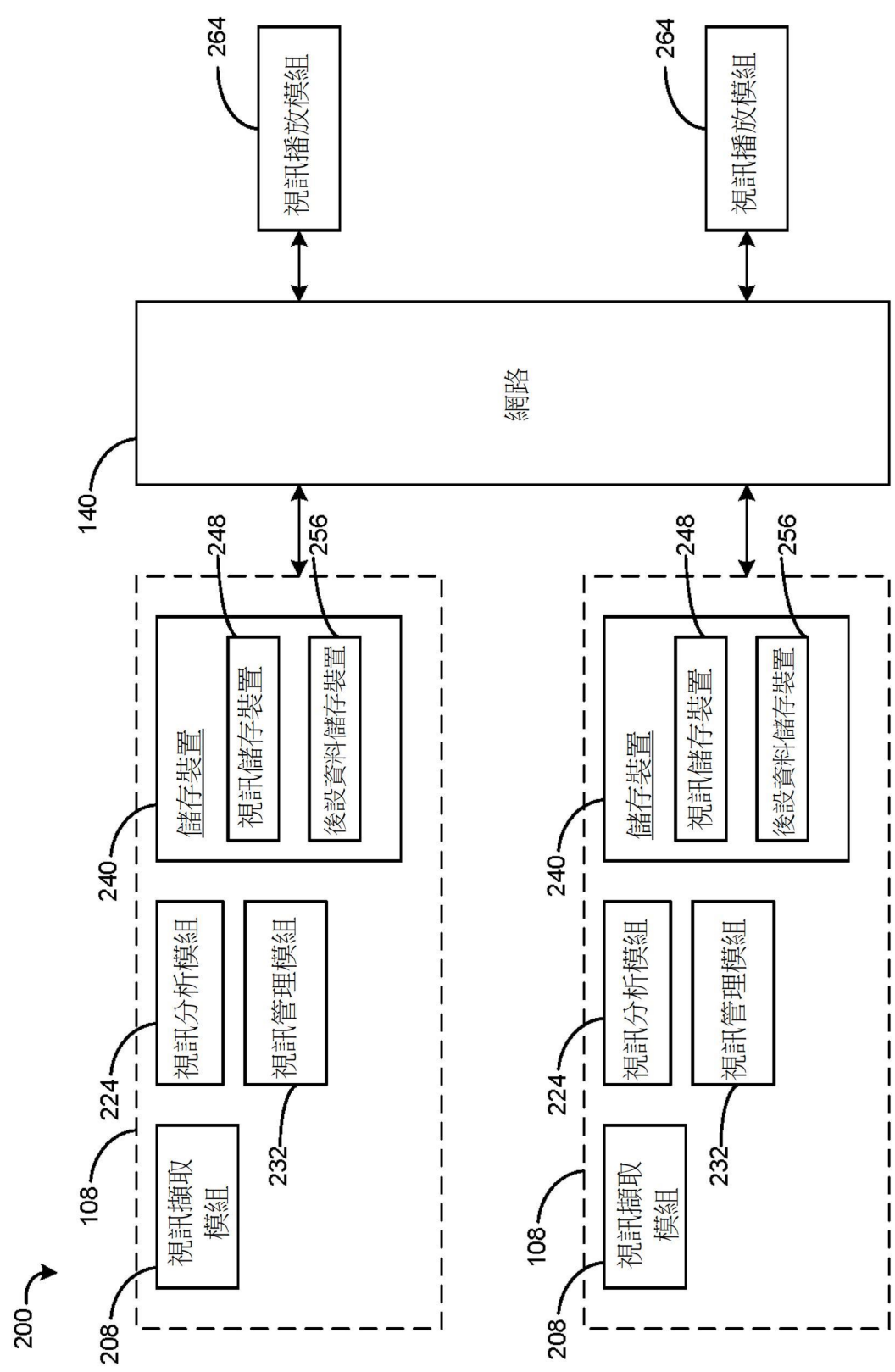
【發明圖式】



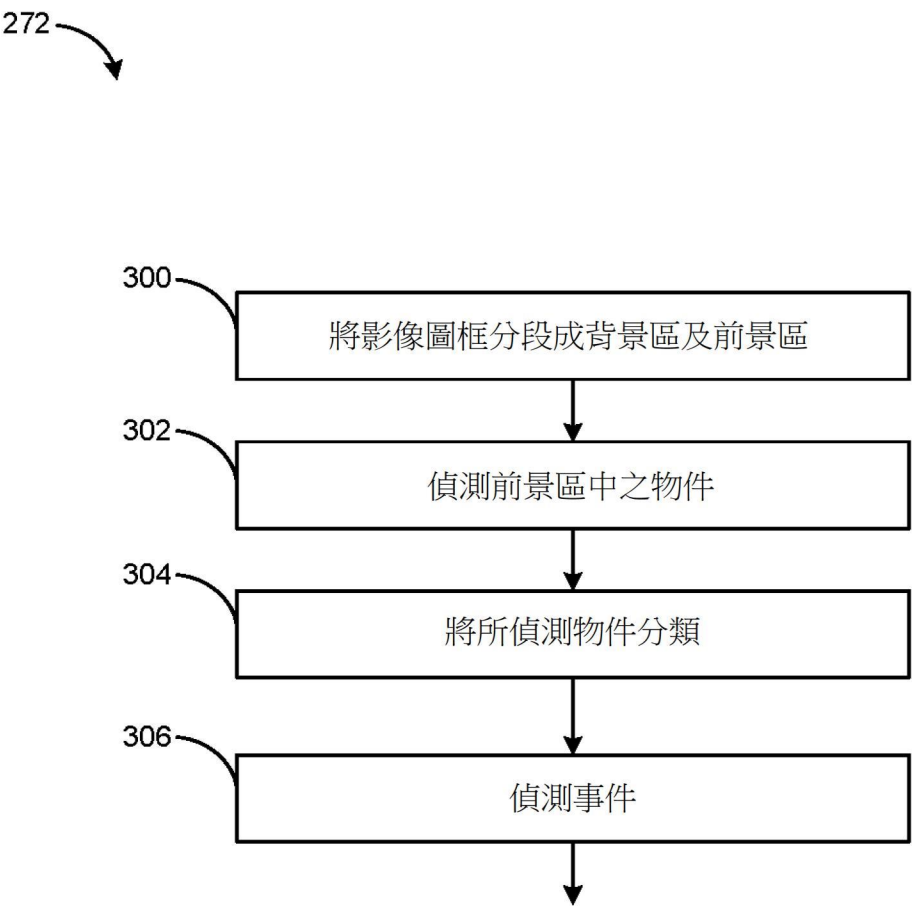
【圖 1A】



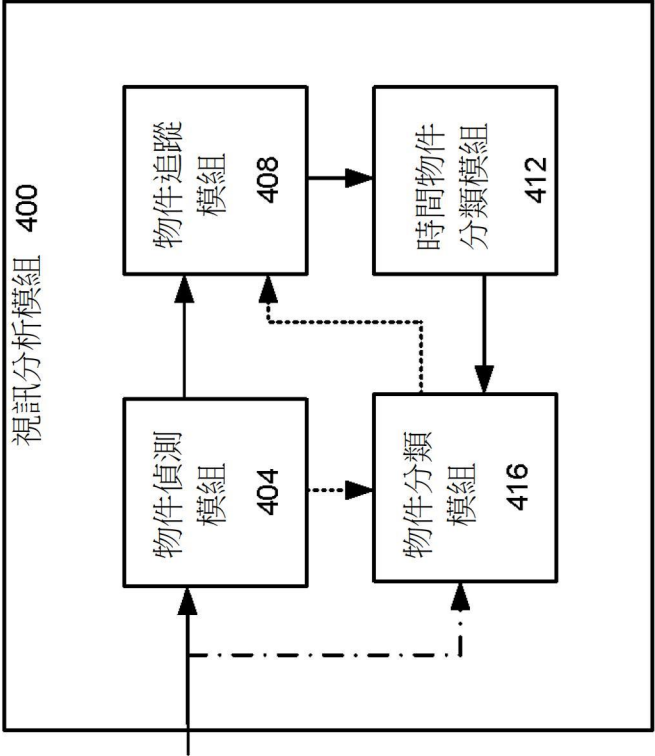
【圖 1B】



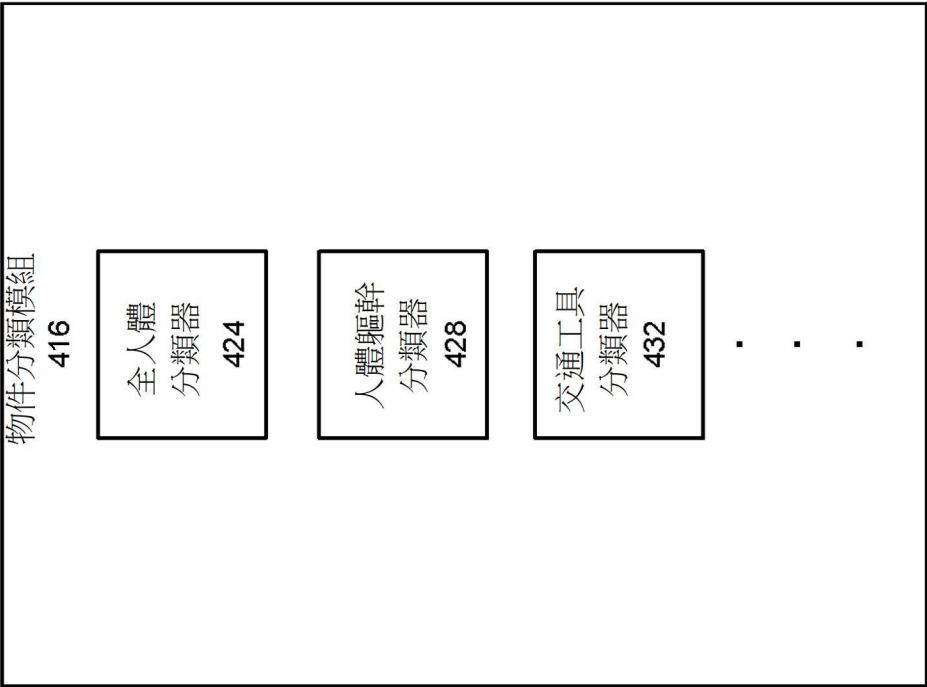
【圖 1C】



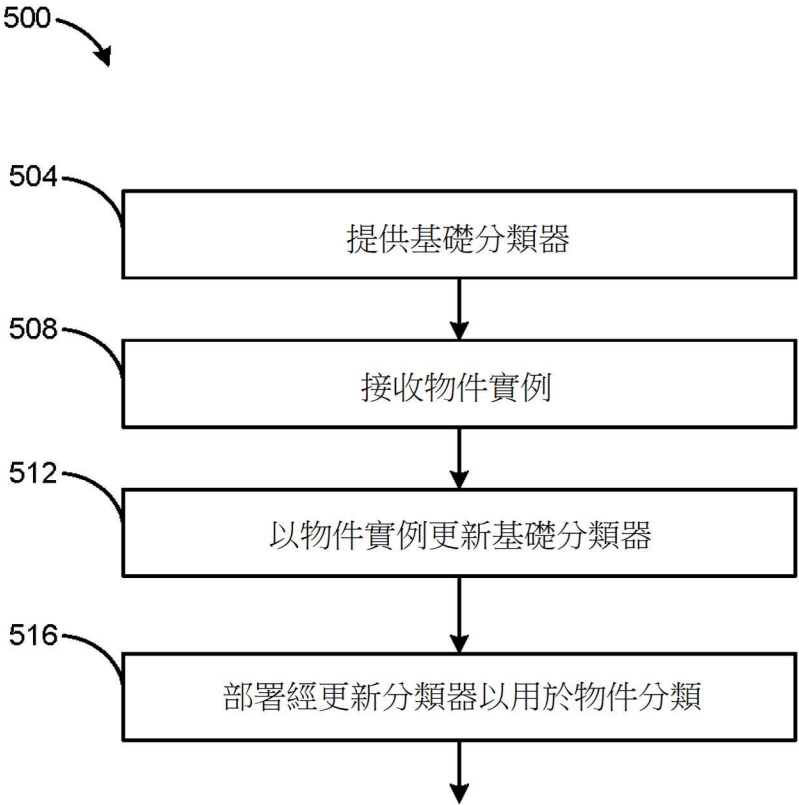
【圖2】



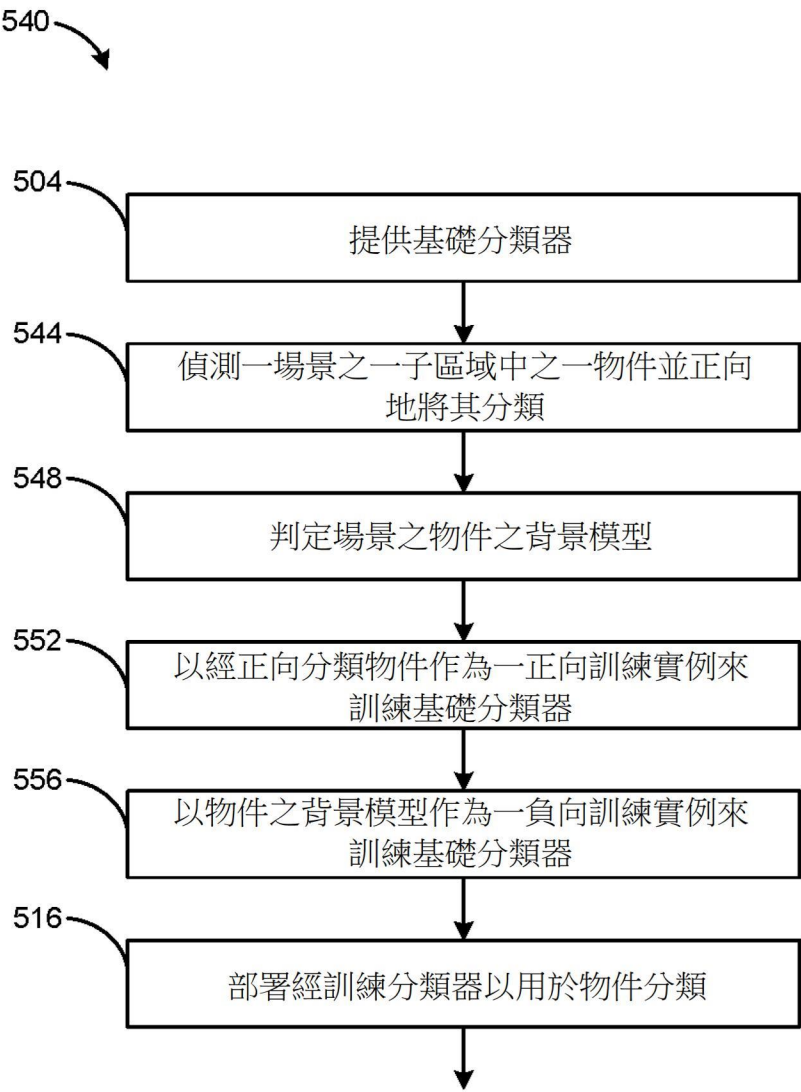
【圖 3A】



【圖 3B】



【圖4】



【圖5】



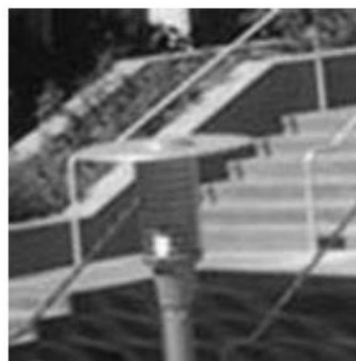
【圖 6A】



【圖 6B】



【圖 6C】



【圖 6D】



【圖 6E】



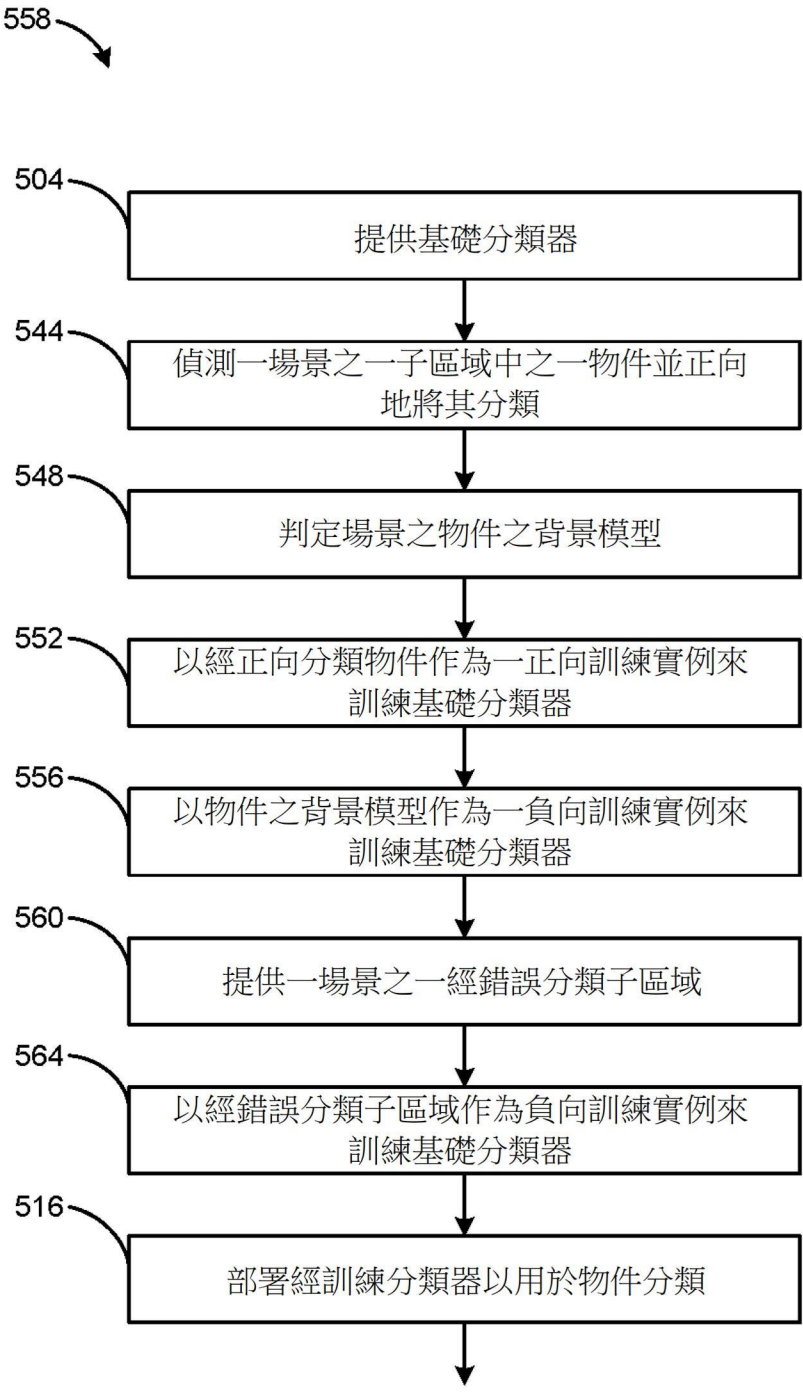
【圖 6F】



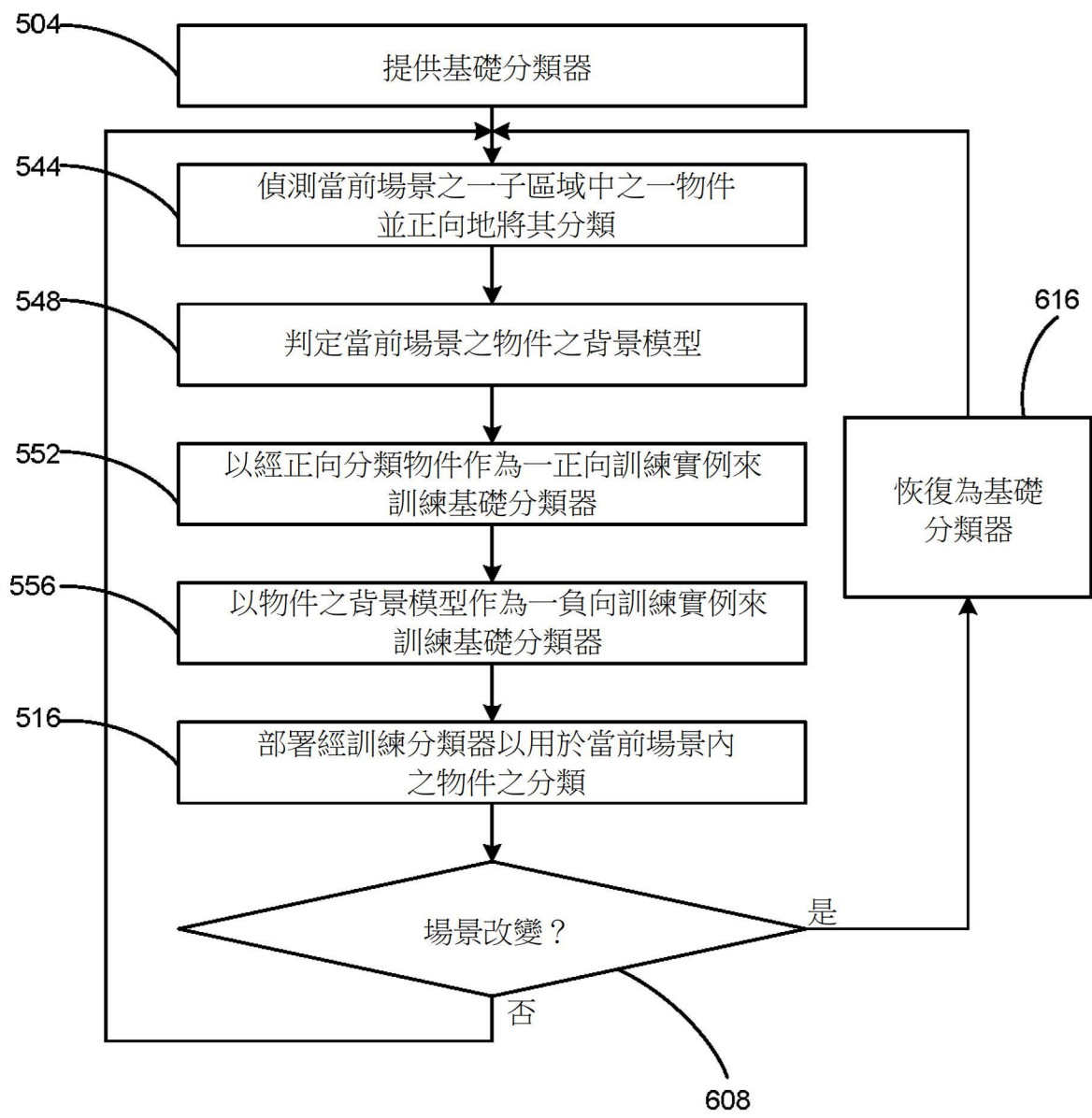
【圖 7A】



【圖 7B】



【圖8】



【圖9】