

图 1

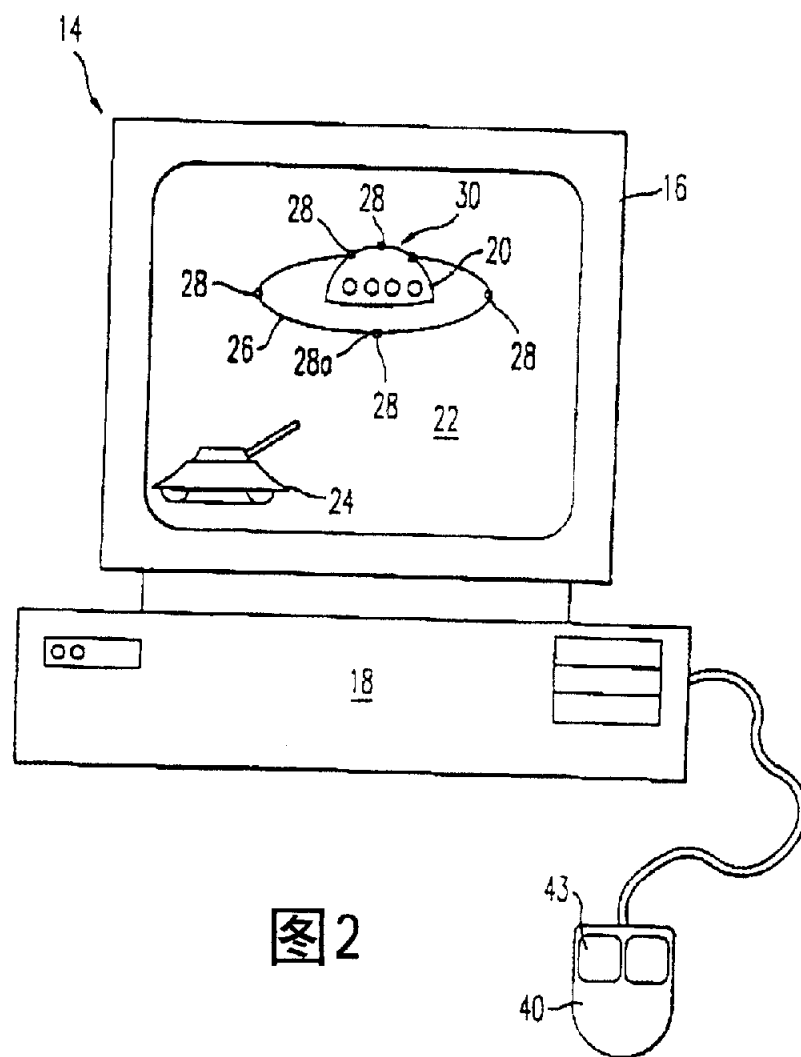


图2

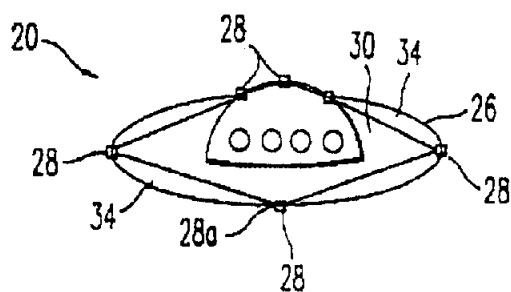


图 3

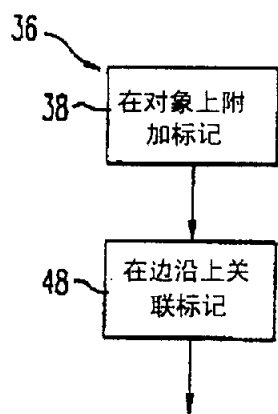


图 5

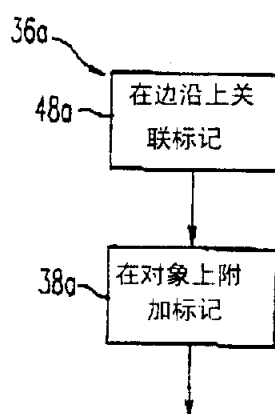


图 6

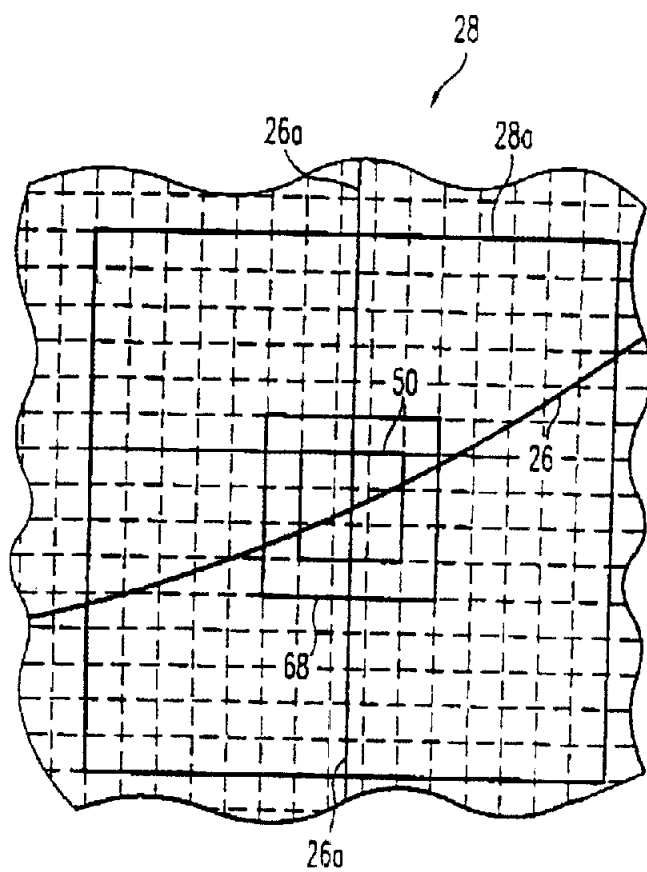
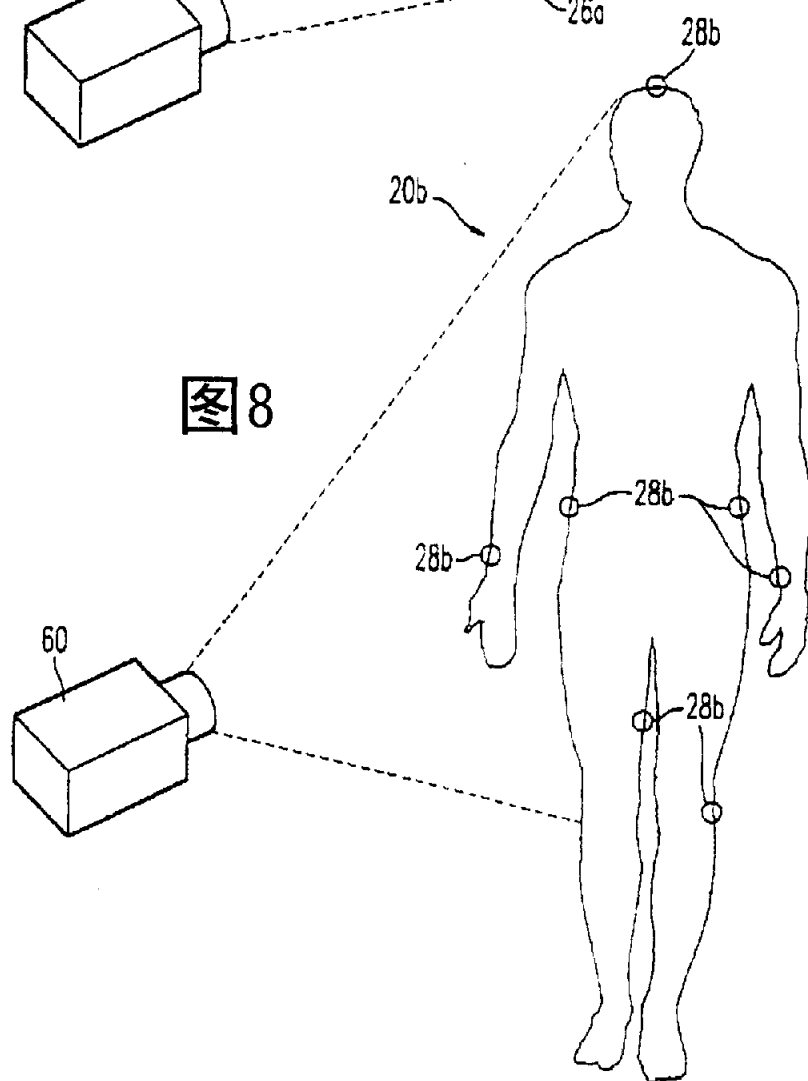
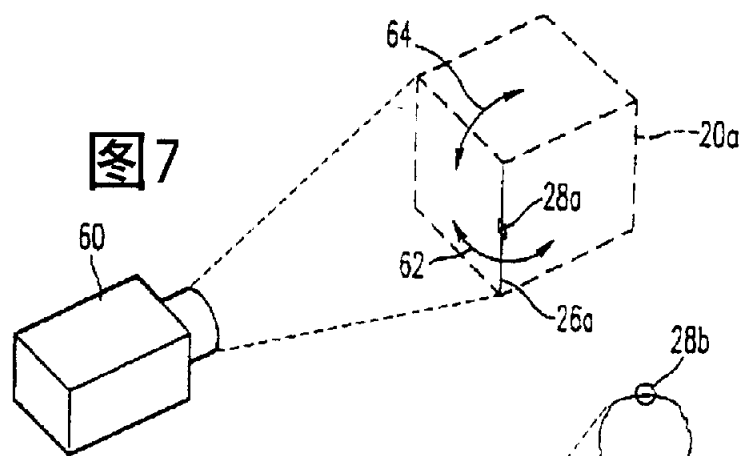


图 4



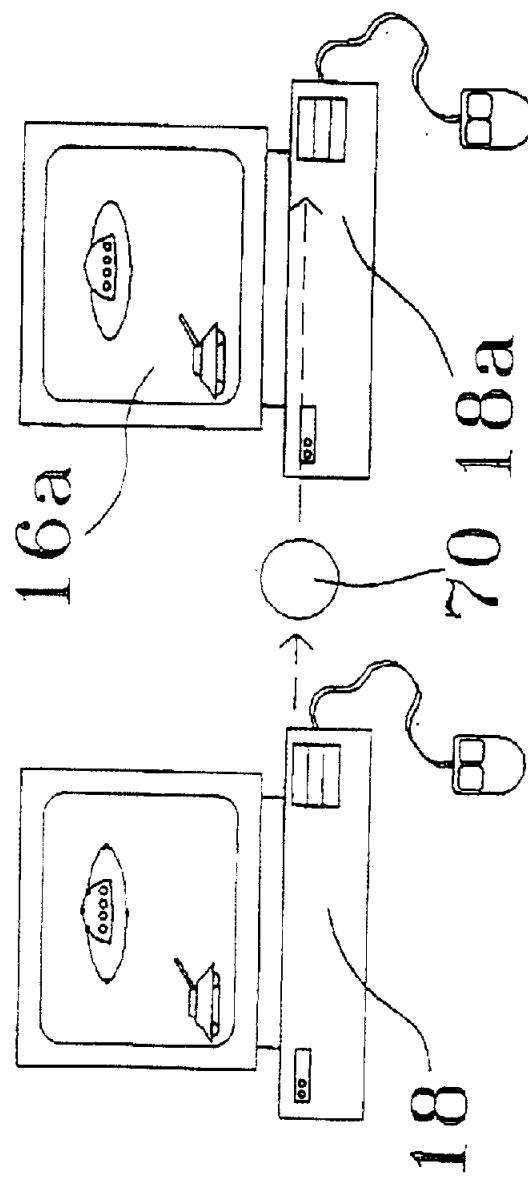


图9



[12] 发明专利申请公开说明书

[21] 申请号 96194232.0

[43]公开日 1998年6月24日

[11] 公开号 CN 1185850A

[22]申请日 96.5.10

[30]优先权

[32]95.5.10 [33]US[31]08/435,439

[86]国际申请 PCT/US96/06542 96.5.10

[87]国际公布 WO96/36007 英 96.11.14

[85]进入国家阶段日期 97.11.28

[71]申请人 魔环媒体公司

地址 美国加利福尼亚州

[72]发明人 安东尼·J·伊莎多拉-巴雷卡

戴维·R·恩伯森 马克·I·希梅斯坦

唐纳德·E·迈耶

[74]专利代理机构 永新专利商标代理有限公司

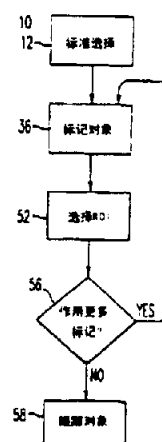
代理人 蹇 炜

权利要求书 3 页 说明书 15 页 附图页数 6 页

[54]发明名称 活动视频图象中的对象识别

[57]摘要

一种识别与跟踪视频图象(14)的对象(20)的活动对象识别方法(10),使对象(20)对于交互式计算机/用户界面(70)能作为热点(30)工作。多个标记(28)定义对象(20)的边沿(26),而多条热点边界(32)定义热点(30),使得热点(30)与对象(20)大致重合。可选用物理标记(28b)来定位标记(28)。对边沿(26)的消失的灵敏度是可以按照子标记(68)对对应的标记(28)的相对大小调整的。



权 利 要 求 书

1、一种识别数字化活动视频图象内的对象的方法，包括：

a) 在活动视频图象上放置若干标记，使所述标记大致勾划出该对象；

b) 令所述标记随该活动视频图象前进而移动，使得即使对象在活动视频图象上中移动与改变形状时所述标记也继续大致勾划该对象。

2、权利要求1的方法，其中：

a) 所述标记包括像素的一个二维阵列；以及

b) 将所述标记初始放置在活动视频图象上使所述标记包含该对象的一条边沿。

3、权利要求1的方法，其中：

a) 用包含多个转发机的装置将所述标记初始放置在对象上，使转发机定义该对象的边沿。

4、权利要求1的方法，其中：

a) 将所述标记的位置存储在计算机中，使计算机能识别热点区，该热点区为所述标记所定义的视频图象部分。

5、权利要求1的方法，其中：

a) 当起始用户为各所述标记在所要求的位置上摁按钮时，便将各所述标记放置在该对象上。

6、权利要求1的方法，其中：

a) 各所述标记在其中包括一个较小的子标记；以及

b) 当一条边沿落在子标记内且进一步当该边沿不落在灵敏区内时，将至少一个标记内的边沿报告丢失的，灵敏区为不存子标记内的标记部分。

7、一种在计算机的显示屏幕上建立热点的方法，包括：

a) 定位多个标记，使所述标记大致上定义数字化活动视频图象中的对象的形状；



b) 当对象在活动视频图象内移动时, 令所述标记在对象上保持它们的相对位置; 以及

c) 记录所述标记的位置, 将由所述标记限定的活动视频图象部分定义为热点。

8、权利要求7 的方法, 其中:

a) 至少一个所述标记位于对象的边沿上。

9、权利要求8 的方法, 其中:

a) 将标记初始放置在对象上, 使边沿落在标记内。

10、权利要求8 的方法, 其中:

a) 当边沿在活动视频图象中移动时, 将标记移动而使边沿保持定中在标记内。

11、权利要求8 的方法, 其中:

a) 在标记内定义一个感兴趣的区, 使得该感兴趣的区是活动视频图象中标记与对象重合的部分; 以及

b) 当感兴趣的区的彩色按照预设的极限改变时, 便报告标记丢失。

12、权利要求8 的方法, 其中:

a) 在标记内定义一个感兴趣区, 使该感兴趣区为活动视频图象内标记与对象重合的部分; 以及

b) 当感兴趣区的亮度按照预设的极限改变时, 便报告标记丢失。

13、权利要求8 的方法, 其中:

a) 当边沿在标记内成为透视缩短而边沿不再延伸到预定义的灵敏区中时, 便丢失了标记。

14、权利要求13 的计算机/ 用户界面, 其中:

a) 灵敏区为活动视频图象落入标记内并在子标记外的部分, 其中该子标记为标记的一个较小全封闭子集。

15、权利要求14 的方法, 其中:



a) 灵敏区至少与边沿检测框一样大, 边沿检测框为用于检测边沿的一个像素阵列。

1 6 、权利要求7 的方法, 其中:

a) 当执行用于在计算机的显示屏上建立热点的方法时, 初始用户能看见这些标记; 以及

b) 最终用户看不见这些标记, 使得最终用户通常不能区别对象与计算机可识别的热点。

1 7 、一种在活动视频图象内跟踪对象的方法, 包括:

a) 选择能用来初始定义对象的标准;

b) 根据在步骤a 中选择的标准将标记放置在对象上;

c) 在标记内选择一个感兴趣区, 该感兴趣区为标记与对象重合的部分; 以及

d) 重复步骤b 与c 直到用多个标记定义的一个热点大致上与对象重合为止。

1 8 、权利要求1 7 的方法, 还包括:

a) 在步骤d 后面, 令多个标记的各个根据在步骤a 中选择的标准去跟踪对象。

1 9 、权利要求1 7 的方法, 其中:

a) 当一个热点边界不再大致上定义对象时, 初始用户能重新启动本创造性方法, 热点边界为互连热点的线, 起始用户能看见该线叠加在活动视频图象上。

活动视频图象中的对象识别

本发明一般涉及计算机视频图象操作领域。而更具体地涉及用于识别视频图象内所关心的对象从而即使对象在图象内移动也能保持对象的识别的改进的装置与方法。

静止画面及活动视频图象的数字化视频图象操作为面向将“多媒体”引入我们生活中许多方面的当前趋势的重要方面，也是诸如制作动画等较传统的努力的现代方向。具有与本发明同一发明人的共有未决美国专利申请号0 8 / 1 4 6 , 9 6 4 教导了将传统的“动画”图象转换成计算机/ 用户界面装置的方法。在实行该先有发明的方法中，必须在视频图象中识别所关心的特定对象。如在上面引用的公开中所讨论的，这种识别可以是相当费力的，并曾预言有可能在未来开发出将这种劳动中的某些从人类操作员转移给计算机的方法。公开了指定由活动图象内的对象构成的“热点”为“…通过观看各关键帧及至少为之研制出更自动化的系统之前手工指定要在关键帧中作为热点指定的感兴趣的对象或东西（如果有的话）”（在这一引文中已删除了关于前面的指示的参照号码）。本申请指向用于自动化这种对象的识别及在时间上保持这种识别的方法与装置。虽然本创造性方法旨在结合交互式计算机界面系统制作使用，但并不限于这种应用。

动画或其它专门制作的活动的视频图象中的对象能比较容易地识别，因为这种对象能按照特定的容易区别的标准（诸如颜色之类）来制作，或者甚至在制作该对象时能使该对象的象素位置成为在计算机内描述该对象的数据的一部分。然而，在并非专门产生或制作的现场动作图象中的对象便不能这样容易地隔离。

识别视频图象中的对象，诸如按照计算机可以理解的标准定义的对象，的先有技术方法包括识别定义该对象的边沿、颜色或颜色模式与/ 或亮度级。对于静止视频图象的识别与/ 或操作，这些方法已相对地有效。例如，能用自动定义其外

侧边沿的已知方法来区别对象，然后在该对象上操作。作为这些操作的示例，可使对象在图象内移动，从图象中消除或改变其颜色、亮度之类。一旦用位图定义了图象，在本发明的上下文中甚至还可以以图符或“热点”的方式使用对象，使得在对象的图象区上摁动按钮能导致计算机启动应答或进一步与最终用户交互作用。然而，应当记住，这种过程不能很好地转移到活动视频图象的领域内。首先，通过为活动图象的所有帧存储全体对象的位图来跟踪活动视频图象内的对象的位置会加重计算机的数据存储器容量的负担并降低计算机的操作速度。第二，即使在单个静止视频图象上工作时执行上述操作所需的用户交互作用与劳动量是可以容忍的，试图为一个活动图象每秒重复三十次左右这一操作会很快显露这一方法超出了可实现的范围。

用来分离活动视频图象中的对象的一种采纳的方法是基于图象的部分的颜色的。熟悉本技术的人员是熟知“兰色屏幕”法的，其中能够有选择地从图象中消除具有特定颜色（通常但并非必要是兰色）的图象的部分。这一技术在数字图象操作到来之前已用在电视中，但已发现应用在数字图象操作领域中时也工作得很好。虽然这一方法对其意想的目的工作得很好，但它通常只能成功地将背景诸如与前台对象区分。再者，它需要专门设定所关心的对象必须是相对于特殊的背景颜色原始“拍摄”的（意指拍摄、录在录象带上等）。更重要的是，虽然背景能与前台对象区分使得计算机能计算对象的位置以便在其上执行操作（诸如将对象叠加在不同的背景上），但对于计算机能使用对象来作为计算机/用户交互作用的手段来说，通常不能充分地识别不同的对象。再者，即使能够在兰色背景之类前方拍摄视频场景的特殊情况中，以及甚至在前台只有一个单一的对象而不存在对象之间的混淆的寻常的情况中，这种先有技术解决方法也不能解决巨大的数据存储器需求及计算资源枯竭问题，如上所述。

与更人为地引入的背景“兰色屏幕”法相比，识别视频图象内的边沿或复杂颜色模式的方法对于分离视频图象的特定“现实世界”部分更有效。然而，这种边沿或颜色识别法通常需要相对复杂的计算机分析，因此不适合于实时图象跟踪，



至少除非将大量昂贵的计算能力专用在这种跟踪上。即使在任务可以利用实际上无限制的计算机资源量时，也已证明试图按照这种复杂的标准来跟踪视频图象内的活动对象也是令人不快地复杂的。当在图象中要跟踪一个以上对象时，或者当对象在图象内快速地移动与/或改变相对形状时，与这些方法相关联的问题便恶化了。

本发明人已注意到，为了实现他们自己的产品，若干计算机产生的卓越制造商具有对用于识别视频图象内的对象而使计算机能跟踪这些对象的更好地装置与/或方法的需求。然而，尽管这些公司中一些具有巨大的研究预算以及许多有经验的研究人员，它们还是转向本发明人寻求对这一已知问题的解决方案。

据本发明人所知，在先有技术中不存在用于计算机跟踪与操作的快速与容易地识别活动视频图象中的对象的廉价与容易实现并且可靠的可工作的方法。所有先有技术方法或者极为费事与/或需要异常大的计算力量来实现（或者甚至更坏，需要最终用户的极大计算力量来利用这些方法的产品）或者不可靠地识别对象而使计算机不能在不丢失对象或将它们与背景或图象中的其它对象混淆地跟踪视频图象中的对象。

从而，本发明的目的为提供容易地识别数字化活动视频图象中的对象而使计算机能跟踪这些对象的方法。

本发明的另一目的为提供识别数字化活动视频图象中的对象而只需要最少的操作员介入与劳动的方法。

本发明的又另一目的为提供不需要大量计算能力来实现的识别数字化活动视频图象中的对象的方法。

本发明的又一目的为提供计算机能使用最少量的数据存储器资源来存储识别标准的方式识别对象的识别数字化活动视频图象中的对象的方法。

本发明的又另一目的为提供在对象运动通过图象时计算机能容易地跟踪对象并且不过分增加计算机资源的负担的识别数字化活动视频图象中的对象的方法。

本发明的又另一目的为提供使对象能以计算机图符之类的方式作为计算机/



用户交互作用的通路的识别数字化活动视频图象中的对象的方法。

简言之，本发明的较佳实施例为识别数字化活动视频图象中的对象的自动化方法，即使在对象改变形状或在视频图象中移动时也能不过分增加计算机的负担而使之能容易地识别这些图象。本创造性活动对象识别方法首先在活动视频图象的一个单一的帧内识别一个对象。初始识别能用众多的起始用户输入或者用更自动化的方法来完成。在本发明的第一较佳实施例中，利用传统的自动化软件方法于初始对象识别。在本发明的同样较佳替代实施例中，利用软件与硬件部件的独一无二的组合于初始对象识别。

虽然按照本发明，初始对象识别是根据已知的方法或已知方法与装置的新的组合完成的，本发明的主要创造性方面在于初始识别后面跟随着只在其中保留了可靠地跟踪对象所必需的数据的转换操作。因此，当对象在活动视频图象中逐帧移动与/ 或改变形状时，只需极少量重新计算来跟踪对象。

关于本发明，应理解计算机既在识别视频图象中的对象的创造性方法中扮演一部分，并且计算机（更为可能的是，用于原来识别对象的以外的不同计算机）将被用来最终使用本发明的最终产品。这一方面，此后将引用“起始用户”，它们是使用本创造性方法来建立计算机/ 用户界面或这里将要公开的创造性方法的其它适当应用的用户。对“最终用户”的引用便是指这些人而言，他们不直接使用本创造性方法，而是使用由起始用户按照本创造性方法产生的计算机/ 用户界面装置。

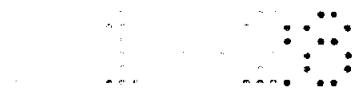
本发明的优点为不需要昂贵的高性能计算机来实现。

本发明的另一优点为不需要昂贵的高性能计算机来使用本方法的最终产品。

本发明的又一优点为不需要大量的用户输入。

本发明的又另一优点为使计算机能可靠地跟踪数字化活动图象中的对象，清楚地将对象与背景及其它对象区别开，从而使计算机与/ 或与计算机联合行动的计算机用户能够在感兴趣的对象上操作。

鉴于实现本发明的最佳当前已知模式的描述及这里所描述的及若干附图中所



例示的较佳实施例的工业应用，对于熟本技术的人员，本发明的这些与其它目的与优点将是显而易见的。

图1 为描述按照本发明的活动对象识别方法的流程图；

图2 为诸如按照本创造性方法在其上操作的视频图象的描述；

图3 为图2 的感兴趣的对象的详细视图，示出对象对热点的关系；

图4 为按照本发明的热点定义标记的详细视图；

图5 为图1 的标记对象操作的更详细流程图；

图6 为展示图1 的标记对象的替代子步骤序列的更详细流程图；以及

图7 为展示简化的对象对摄象机的关系的图形表示。

实现本发明的最佳当前已知模式为主要通过利用计算机建立的活动对象识别方法。虽然该方法实质上可用在其中必须跟踪活动视频图象中的对象，但该创造性活动对象识别方法的突出预期用途是在制造交互式计算机/ 用户界面系统中。

本创造性活动对象识别方法在图1 中的流程图中描述，并在其中用总的参照字符1 0 指示。活动对象识别方法的第一步为“标准选择”操作。

图2 为正显示在计算机1 8 的显示屏1 6 上的视频图象1 4 的图形表示。视频图象1 4 为已按照传统方法数字化而使得视频图象1 4 能存储在计算机1 8 并从而显示在显示屏1 6 上的活动视频图象的一个单一的“帧”。应指出，虽然熟悉本技术的人员认识到在动画电影片、电视信号与数字化视频图象之间在格式上存在实质性差别，但关于这些不同的媒体的许多术语是可以在三者之间传递的。因此，虽然图1 的视频图象1 4 是数字化活动视频图象的一个元素，将视频图象1 4 作为照相编码的活动画面图象的一“帧”来讨论是方便的。熟悉相关技术的人员容易理解数字化视频图象1 4 与概念上对应的动画“帧”之间的关系。本技术中常用这一对照方法，而视频“帧”的意义可认为与单一的视频图象1 4 对应，虽然可以不存在帧的实际物理实施例。

在图2 的视图中，可以看到对象2 0 描绘在背景2 2 上。在这一意义上，背景2 2 是除所关心的对象2 0 以外的全部视频图象1 4 。背景2 2 可包含不可名



状的且通常无法定义的部分，如图2 的示例中的其它东西2 4，该其它东西2 4 是背景2 2 的潜域可定义部分，可以也可以不将它分开定义为附加的感兴趣的对象。

用于识别视频图象1 4 中的对象2 0 从而使计算机1 8 能在其上执行诸如改变对象2 0 的颜色或在视频图象1 4 中删除或移动它之类的某些操作的方法在本技术中是已知的。在本发明的最佳当前已知实施例1 0 中，利用了称作S o b e l 算法 (G o n z o l e z 与W o o d s 在“数字图象处理”中详细讨论，A d d i s o n - W e s l e y 出版社，1 9 9 2，P 1 9 7 - 2 0 1 及4 1 6 - 4 2 0) 的著名与广泛实行的边沿检测技术来检测对象2 0 的边沿2 6。根据G o n z o l e z 与W o o d s，边沿2 6 为图象的两个区之间的边界 (在图2 的当前示例中，这两个区为对象2 0 与背景2 2)，这两个区相对地具有不同的亮度特征，然而除了或替代用亮度特征之外必然也可以想到根据颜色特征等来定义边沿2 6。

图2 中描绘了多个 (在图2 的实例中为6 个) 标记2 8。对象2 0 上的一个或多个标记2 8 的组合定义一个热点3 0。即对于最终用户，热点3 0 与对象2 0 在感性上实质上通常是一个且是同一个。然而，如这里所定义的，只将对象2 0 看作是人类观看者可定义的实体的视频图象1 4 上的区域，而热点3 0 则是定义成计算机1 8 能识别与跟踪的热点3 0，即使热点3 0 在视频图象1 4 中改变位置与/ 或形状也一样能识别与跟踪。

图3 为图2 的对象2 0 的更详细的视图。在图3 的视图中，能够看出热点3 0 是用标记2 8 定义的，这便建立了热点的边界3 2。这便可看出，热点3 0 与对象2 0 大致上但不是精确地占据视频图象1 4 (图1) 的同一部分。在主要意向的应用中这没有实际意义，因为在对象2 0 边界内但尚在热点2 0 的范围外的任何未定义区3 4 将在对象2 0 的外边界上。试图在对象上摁按钮的最终用户将自然地趋向于在对象2 0 的中心附近摁按钮，因而也靠近热点3 0 的中心。即使最终用户看不见也无法知道热点边界3 2 的确切位置，最终用户也会本能地不期

望来自对象2 0 的边沿附近的未定义区3 4 中的摁按钮的响应。能够理解，正确地分布的标记2 8 的量越大，热点3 0 与对象2 0 之间的互相关联也越大。

图4 为图2 的代表性标记2 5 a 的详图。如前所述，边沿2 6 是按照本技术中熟知的原理与方法定义的。具体地，在本发明的最佳当前已知实施例1 0 的实例中，采用了广泛实行的S o b e l 算法来识别边沿2 6 。应指出，边沿识别进程通常不加区别地识别边沿，从而不仅对象2 0 的边沿2 6 ，并且可能存在视频图象1 4 （图2 ）中的任何其它不同的边沿（未具体标识的）也被这样识别出。

再参见图1 的流程图，如这里关于本发明的最佳当前已知实施例1 0 的实例所描述的，能够看出“标准选择”操作1 2 为选择使用S o b e l 算法根据亮度标准检测边沿（包含特别感兴趣的边沿2 8 ）。在活动对象识别方法的某些应用中，“标准选择”操作1 2 有可能对于该应用的所有使用是固定不变的，从而在每一次执行该创造性活动对象识别方法1 0 时，起始用户将不执行“标准选择”操作1 2 。

再参见图1 的流程图，能看出在“标准选择”操作1 2 后面执行“标记对象”操作3 6 。如此后将更详细地描述的，在本发明的最佳当前已知实施例1 0 中，虽然在本发明的范围内将来有可能研制出某些装置或方法来实质上完全自动化“标记对象”操作3 6 ，或者至少为了减少对用户干预的需求而进一步帮助起始用户，但当前需要起始用户的干预来执行“标记对象”操作3 6 。

在本发明的最佳当前已知实施例1 0 中，用户在“标记对象”操作3 6 中通过在对象的边沿2 6 上摁按钮而将标记2 8 放置在感兴趣的对象2 6 的边界周围。图6 为图1 的“标记对象”操作3 6 的更详细流程图。在图5 的视图中，能够看出“在对象上附加标记”子步骤3 8 中“在边沿上关联标记”子步骤4 0 前面。按照本发明的最佳当前已知实施例1 0 ，起始用户将在对象2 0 的边沿2 6 上“摁按钮”，诸如利用鼠标器4 0 （图1 ）将光标4 2 移动到边沿2 6 上，然后起始用户摁下鼠标器4 0 的按钮4 4 。应记住，起始用户在进程的这一点上必须用视觉看到边沿2 6 的存在。



再参见图4 的视图，其中标识了多个像素4 4 。熟悉本技术的人员将理解显示屏1 8 上具有许多像素4 4 ，其确切数量将是显示屏1 8 本身以及控制视频图象1 4 的硬件与软件的特征的函数。为了便于描述本创造性活动对象识别方法1 0 ，将像素4 4 示出在网格上，使得像素4 4 呈现为大致上方形的，虽然熟悉本技术的人员将理解实际的物理像素（未示出）可以是圆的、矩形的或其它形状的。

在本实例中，如果起始用户在选择点4 6 上摁按钮（选择点4 6 是在起始用户放置光标4 2 （图2 ）并摁下鼠标按钮4 4 时建立的），便将标记2 8 a 以其中心作为选择点4 放置。在图4 的示例中，标记2 8 a 的大小是9 ×9 个像素3 6 ，该大小的意义将在下面更详细地讨论。

再参见图5 ，上述“在对象上附加标记”操作3 8 后面跟随“在边沿上关联标记”操作4 8 。按照本发明的最佳当前已知实施例1 0 ，在“在边沿上关联标记”操作4 8 中，在标记2 8 a 内自动检测边沿2 6 ，并在边沿2 6 及在标记2 8 a 的中心上建立边沿框5 0 。如前所述，为了展示本发明的最佳当前已知实施例1 0 的目的，在“标准选择”操作1 2 （图1 ）中选择了S o b e l 算法作为边沿检测手段。熟悉S o b e l 算法的人员将注意到边沿检测进程中通常采用3 ×3 的框，因此在采用S o b e l 算法时，对于边沿框5 0 的大小为三个像素4 4 乘三个像素4 4 是方便的。令标记2 8 a 的大小为n ×n 个像素4 4 也是方便的，其中n 为大于1 的整数，如图4 的9 ×9 大小的标记2 8 a 所示。

当然，可以采用修正的S o b e l 算法而使边沿框5 0 具有另一大小，或者作为替代可采用完全不同的边沿检测方法，从而按照这里详述的本发明的其它参数，边沿框5 0 实际上可以是任何大小。

应指出，图4 的实例中，如此前所述，起始用户先前已在边沿2 6 上摁过按钮而使选择点4 6 落在边沿2 6 上。在实际实践中，起始用户对边沿2 的位置的估计有可能不够精确从而使选择点4 6 的初始位置不能精确地落在边沿4 6 上。这一情况是比较容易纠正的，纠正方法如下：当在“在对象上附加标记”子步骤

3 8 中放置了标记2 8 a , 并随后在“在边沿上关联标记”子步骤4 8 中检测到边沿2 6 并将边沿框5 0 定中在边沿2 6 上时, 如果边沿框5 0 未定中在标记2 8 a 上, 则计算机1 8 自动地重新定位标记2 8 a 而使边沿框5 0 定中在标记2 8 a 上。设想过可以采用替代的方法来处理在标记2 8 a 中没有检测到边沿2 6 的部分的情况。例如, 可知起始用户“未检测到边沿”的情况并通知他再试验。作为替代, 可将边沿的搜索扩展到标记2 8 a 以外预定数目的像素4 0 上。如果在这一扩展的搜索中检测到边沿2 6 , 便可相应地自动调整边沿框5 0 与标记2 8 a 的位置。如果即使在这一扩展搜中仍未检测到边沿2 6 , 则可以通知起始用户重复“在对象上附加标记”操作3 8 。

图6 为当前不作为本发明的最佳当前已知实施例1 0 的操作次序提出的替用“标记对象”操作3 6 a 的流程图, 而只是展示了在某些意想不到的环境中可能有用的最佳当前已知实施例1 0 的变型。从图6 的流程图能够看出, 与图5 的同名子步骤3 8 与4 8 相比, 将替代的“在对象上附加标记”子步骤3 8 a 与替代的“在边沿上关联标记”子步骤的次序颠倒。这一次序颠倒要求此前已讨论过的“在对象上附加标记”子步骤3 8 与“在边沿上关联标记”子步骤4 8 的某些修改。

如上所述, 在“在边沿上关联标记”操作4 8 中, 边沿检测(在本例中根据 Sobel 算法)通常出现在标记2 8 a 内。在图6 的替代实施例中, 在完成了“标准选择”操作1 2 时, 在整个视频图象1 4 上运行边沿检测进程(图1)使得边沿2 6 (以及这里没有具体描绘出的视频图象1 4 内的任何其它边沿)突出, 从而起始用户能看见边沿2 6 。通过比较, 在前面描述的“在对象上附加标记”操作3 8 中, 起始用户可估计计算机1 8 有可能在何处理检测到边沿2 6 , 而在本替用实例中, 在替代的“在边沿上关联标记”子步骤4 8 a 一开始, 起始用户便知道了计算机1 8 所确定的边沿2 6 的精确位置, 而使得起始用户能将标记2 8 a 精确地放置在边沿2 6 上。起始用户在替代的“在边沿上关联标记”操作4 8 a 中将标记2 8 a 放置在边沿2 6 上之后, 计算机1 8 在替代的“在对象上附



加标记”子步骤3 8 a 中定中标记2 8 a 便是相对简单的事了，从而边沿框5 0（根据替代的“标记对象”操作3 8 a，它已经与边沿2 6 的一部分重合）便对称地布列在边沿框5 0 四周。

“标记对象”后面（或者上述替代的“标记对象”操作3 8 a 后面）为“选择感兴趣的区域”（“选择ROI”）操作5 2。在“选择ROI”操作5 2 中，起始用户只是在对象2 0 所在的边沿2 6 的一侧摁按钮来指定感兴趣的区域（“ROI”）5 4，ROI 5 4 便是视频图象1 4 中标记2 8 a 与对象2 0 重合的那一部分。

在“选择ROI”操作5 2 后面，在“放置更多标记？”判定操作5 6 中，给予起始用户选择是否想要在对象2 0 上放置更多的标记2 8。如果想要放置更多的标记2 8，则本创造性进程返回到“标记对象”操作3 8。如果对“放置更多标记”判定操作5 6 的回答为否“N”），则进程继续到“跟踪对象”操作2 8。

应指出，已在前面提出的创造性活动对象识别方法1 0 的操作是在活动视频图象1 4（图2）的单一“帧”上进行的。现在开始讨论当视频图象1 4 通过时间改变时本创造性过程如何继续进行，熟悉本技术的人员会认识到在数字化现场动作视频制品中，由于各种噪声因素、光照变化、视频图象1 4 中的对象2 0 及其它东西2 4（如果存在）的运动和/或原始捕捉视频图象1 4 的摄象机（未示出）相对于对象2 0 的运动等，诸如边沿2 6 的尺寸、各像素4 4 的彩色值、亮度特征等逐帧属性中各个与全体也能变化。

在讨论通过时间利用标记2 8 跟踪对象2 0 之前，本发明人预期要利用某些参数来判定标记2 8 是否继续精确地表示对象2 0 而使得热点3 0 与对象2 0 大致上重合。根据本发明的预期的最佳当前已知实施例1 0，这些参数为：亮度、颜色及标记灵敏度。

关于亮度，作为“跟踪对象”操作5 8 的一部分，计算机1 8 将计算边沿2 6 两侧标记2 8 内的平均亮度。起始用户可指定作为亮度差的百分比表示的一个



量，在计算机1 8 将各相应的标记2 8 标志为丢失之前，亮度值可在这一百分比之内变化。（“标记丢失”条件的事实与后果将在下面更详细地讨论。）

关于颜色，计算机1 8 在逐个分量（即R G B）的基础上跟踪R O I 5 4 内的象素4 4 的颜色的平均值。彩色参数允许起始用户规定百分比表示的量，在将各相应的标记2 8 标志为丢失之前，这些平均值能在这一百分比之内变化。

关于标记灵敏度：这一参数表示对于在二维空间（视频图象1 4）中跟踪三维对象2 0 所特有的现象，它描述本创造性活动对象识别方法1 0 的自然环境。图7 为展示相对于摄象机6 0 示出的简化的（矩形）对象2 0 a 的视图。只是为了当前的讨论的目的而在图7 的视图中加重了一条垂直边沿2 6 a（只是为了本实例的目的而任意选择的，因为水平的边沿[未特别描绘]或其它这种边沿能用来展示正在说明的问题）。关于图7 的视图，能够看出该简化的对象2 0 a 可以按照第一转动箭头6 2 转动而使得从摄象机6 0 观察时垂直边沿2 6 a 的透视长度保持基本不变。此外，简化的对象2 6 a 可按照第二转动箭头6 4 转动，使垂直边沿2 6 a 的透视长度根据转动量改变。当然，如果简化的对象2 0 a 在第二转动箭头6 4 所指的任一方向上转动整整9 0 度，从摄象机6 0 的视角观看，垂直边沿2 8 a 实质上将变成无限地短。能够理解，取决于该边沿2 6 相对于摄象机6 0 的夹角并进一步取决于如用简化的对象2 0 a（图7）所示出的对象2 0（图2）的实际转动方向，不完全垂直或水平的边沿2 6 将在不同程度上经历这一现象。根据上面的讨论也能理解，摄象机6 0 相对于简化的对象2 0 a 的运动和简化的对象2 0 a 相对于摄象机6 0 的运动一样，也会产生基本上同样的边沿2 6（如用垂直边沿2 6 a 作典型的）的相对伸长与缩短。

现在再参见图4 的视图，灵敏区6 6 是标记2 8 a 不同时在子标记6 8 内的部分。在图4 的示例中，子标记6 8 的大小为5 ×5 个象素4 4。子标记6 8 定中在标记2 8 a 内。实践中，子标记6 8 可以也可以不与边沿框5 0 一样大小并从而与之或不与之完全重合。在图4 的示例中将垂直边沿2 6 a 描绘成水平地定中在代表性标记2 8 a 内的（如同根据此前讨论的“标记对象”操作3 6 或替代



的“标记对象”操作3 6 a 或替代的“标记对象”操作3 6 a 那样)。从图2 的视图能够理解, 当简化的对象2 0 a 如第二转动箭头所指转动时, 便使垂直边沿2 6 a 从垂直方向上移开, 从摄象机6 0 的视角看来垂直边沿2 6 a 的视长度会越来越短。当垂直边沿2 6 a 短到子标记6 8 之内但不在灵敏区6 6 之内时, 便为该特定标记2 8 指示标记丢失条件。如根据从上讨论能够理解的, 响应这一条件的“灵敏度”能够通过调节灵敏区6 6 的大小而加以调节。这便是子标记6 8 相对于标记2 8 越大, 则灵敏区6 6 越小并且在出现边沿2 6 丢失时(如用垂直边沿2 8 a 的示例所示的)指示标记丢失条件越快。(参见图7 的视图会注意到, 代表性标记2 8 a 并不实际叠加在简化的对象2 6 a 上而使得在其上面看见代表性标记。而只是为了表示垂直边沿2 6 a 的一部分也示出在图4 的视图中的目的而在图7 的视图中示出代表性标记2 8 a)。

图8 展示本创造性活动对象识别方法的同样较佳的替代实施例8 1 0 。在图8 的视图中, 能够看到一个复杂的对象2 0 b 上分布有多个物理标记2 8 b 。这些物理标记2 8 b 为能将它们的相对位置转发给摄象机6 0 的转发机, 使得这些相对位置能与其视频图象1 4 (图2) 一起记录。构成物理标记的这种转发机现在已可普遍地购得。它们根据若干技术中任何一种工作, 其中包括无线电三角测量与超声三角测量原理。例如, L o g i t e c h t m 已作为开发人员成套仪器销售能够提供物理标记2 8 b 的非常精确的三维位置信息以及姿态信息(倾斜角、俯仰角与偏航角), 同时不产生可听见的噪声并同时一般抗来自环境噪声的干扰的超声系统。应指出, 本领域中已知在制作电影时将单个传感器附着在对象上(诸如人)用于记录该对象相对于摄影机的位置。然而从前已知与采用的这种对象不够灵敏, 当将它们用在先有技术中采用它们的目的时, 它们也不需要区别这些对象的限制。

当如图8 中所示那样实现物理标记2 8 b 时, 将物理标记的记录的位置传送到软件中便是简单的事件, 从而使计算机1 8 能够根据(软件)标记2 8 定义对象2 0 , 如这里所描述的。从原始记录介质传送理理标记2 8 b 的位置之后而使



计算机1 8 具有视频图象1 4 的这些位置的记录，则物理标记2 8 b 的位置与前面讨论的在软件中原来建立的标记2 8 的位置没有明显的不同。

应指出，不论初始使用的是（软件）标记2 8 或物理标记2 8 a ，将会出现此前未特别提到的某些共同情况。例如，经常会出现随着对象2 0 在视频图象1 4 内转动或移动，当从二维透视图观看时，边沿2 6 会出现跨越。（这便是，例如当对象2 0 整个转动去而使得从静止的视点上看来右边变成左边而左边变成右边时，则在某些中间点上边沿将出现跨越。）实际上这些变化对本创造性活动对象识别方法1 0 提出很少问题或不成问题。当标记2 8 穿过热点边界3 2 而使标记2 8 位于热点边界3 2 内部时，可以简单地暂时忽略它，直到它再一次穿过存在的热点边界3 2 并从热点3 0 内出来为止。

图9 示出用来协助执行本创造性方法的计算机1 8 与在上面使用诸如交互式计算机/ 用户界面7 0 等最终产品的最终用户计算机1 8 a 的关系。交互式计算机/ 用户界面7 0 示出在实施在C D R O M 中的图2 的视图中。当然，交互式计算机/ 用户界面7 0 通常只构成包含它的典型C D R O M 的内容的一个相对地小的部分。在计算机2 8 上产生之后，交互式计算机/ 用户界面7 0 将被装入C D R O M 中（或者作为替代，装入软盘或与最终用户计算机1 8 a 通信的其它装置中）。然后，通常拷贝该C D R O M 并将该C D R O M （或者更经常地，其拷贝）传送给最终用户计算机1 8 a ，在那里将交互式计算机/ 用户界面7 0 引导到其中，然后显示在最终用户计算机1 8 a 的最终用户显示屏1 6 a 上。

再返回到此前数次引用过的“标记丢失”条件的考虑，可注意到单个丢失的标记2 8 通常对于本创造性活动对象识别方法1 0 的延续性不是致命的。如这里所指出的，用于定义热点3 0 的标记2 8 的位置是在计算机1 8 的辅助下由起始用户按照本创造性活动对象识别方法1 0 与视频图象1 4 关联的。如此前所述，起始用户能看到标记2 8 与热点边界3 2 ，而最终用户则不能，因此通常也不能区别对象2 0 与对应的热点3 0 。在边沿2 6 在透视上缩短时，在标记移动得太快而不能按照当前实施的本创造性方法跟踪时，或在一种或另一种噪声屏蔽标记



2 8 时, 标记2 8 便能成为“丢失”的。当告知起始用户标记丢失情况时, 起始用户便能作出判定在热点3 0 与对象2 0 之间是否保持充分的相互关联。如果这仍表现基本上相同, 则“跟踪对象”操作5 8 将被允许继续跟踪对象2 0 及存储其标记2 8 的相关位置。如果作出判定在热点3 0 与对象2 0 之间不再有充分的相互关联, 则起始用户能从中确定需要这样做的视频图象1 4 的运行时间中的时间顺序点上开始重新启动本创造性活动对象识别方法1 0 。

应指出对本创造性方法的修正就是为了减少在初始用户的计算机1 8 中跟踪标记2 8 所需的计算时间设想的。当然, 这不是关于最终用户计算机1 8 a 的考虑, 由于初始计算机1 8 已记录了标记2 8 的位置而最终用户计算机1 8 a 将简单地注意标记2 8 相对于视频图象1 4 的适当位置。然而, 在活动对象识别方法1 0 的实际执行中, 可能有大量的时间包含在跟踪对象操作5 8 中, 该操作是由计算机1 8 执行的。为了减少跟踪标记2 8 所需的计算时间量, 已发现下述方法具有实际价值: 通常, 假设标记2 8 未从其以前的位置移动是方便的。这一假设能很快得到检验, 如果成立, 则节省了花费在重新计算标记2 8 的位置的时间。当证明了上面的假设不正确时, 假定前面的运动方向保持不变是进一步方便的。因此, 计算机1 8 可试验将标记2 8 放置在由该标记1 8 的前面的方向与位移量所确定的位置上, 借此避免在搜索边沿2 6 中必须在较大的区域上移动边沿框5 0 。当然, 如果所有这些假设都证明是不正确的, 计算机1 8 将必须扩大搜索区或者报告前面所述的标记丢失情况。随着本创造性方法的实现与适应各种应用, 可以期望将会研制出与改进本创造性活动对象识别方法1 0 的这些与其它改善。

可以对本发明作出各种修改而不改变其价值或范围。例如, 虽然本创造性方法的当前已知实施例1 0 利用边沿识别来初始定义对象2 8 作为对放置软件标记2 8 的协助, 诸如彩色识别等手段也能同样好地应用。

另一种可能的修改可能是标记的定位的进一步自动化。例如, 当采用了“替代的标记对象”操作3 8 a 而在放置标记2 8 之前识别边沿2 6 时, 可以编写出这样的软件, 在其中起始用户只要在对象内摁按钮, 便能从这样选择的位置上向



外生成多条射线，并将这些射线与边沿2 6 的交点标为标记2 8 。这一方法的进一步改善可允许起始用户从这样产生的标记2 8 中进行选择。

上述全体只是本发明的可得出的实施例的一些示例。熟悉本技术的人员容易看出可以作出无数其它修改与变动而不脱离本发明的精神与范围。从而，以上公开不是为了限制，而应将所附的权利要求书解释为包罗本发明的整个范围。

本创造性活动对象识别方法1 0 旨在广泛地应用 in 多媒体工业中。突出的当前用途为识别与跟踪活动视频图象内的对象，使得能将这些对象好象是图符或类似措施等加以使用，用在最终用户与计算机之间的交互作用上。

实践中，本创造性活动对象识别方法1 0 将被包含在许多更广泛的方法中，其中识别视频图象1 4 内的活动对象2 0 为用于与这些更广泛的方法的产品交互作用的必要或理想手段。原则示例将是在交互式计算机/ 用户界面的生产中，其中最终用户可在对象2 0 上摁按钮来进一步访问信息或启动某些其它动作。例如，当关于这一对象2 0 可以得到进一步的信息时，最终用户便能通过在该对象上摁按钮来访问进一步的信息。

按照本创造性方法，在给定的屏幕上热点3 0 中可能还有热点3 0 与/ 或在一个热点上摁按钮能带出另一屏面或局部屏面，而这些后续屏面还包含其它热点。从而，比先有技术大大地增进了程序或界面的交互性程度。

由于本发明的活动对象识别方法1 0 可以容易地集成进现有的与未来的最终产品生产方法中，可以预期它将在工业中作为对从事计算机软件及软件/ 硬件组合产品生产的人员可利用的新颖与有用的工具受到接纳。由于这些与其它原因，可以预期，本发明的实用性及工业可应用性将是范围广泛与持续时间久远的。