

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

H04N 5/262 (2006.01)

H04N 9/64 (2006.01)



[12] 发明专利申请公布说明书

[21] 申请号 200710099726.5

[43] 公开日 2008年12月3日

[11] 公开号 CN 101316328A

[22] 申请日 2007.5.29

[21] 申请号 200710099726.5

[71] 申请人 中国科学院计算技术研究所

地址 100080 北京市海淀区中关村科学院南路6号

[72] 发明人 刘安安 李锦涛 张勇东 唐胜
宋砚

[74] 专利代理机构 北京泛华伟业知识产权代理有限公司
代理人 高存秀

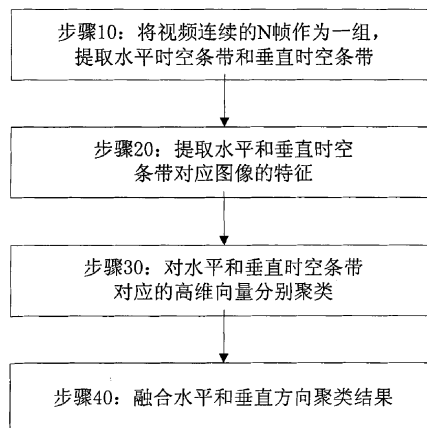
权利要求书3页 说明书8页 附图2页

[54] 发明名称

一种基于时空条带模式分析的新闻主播镜头检测方法

[57] 摘要

本发明公开了一种基于时空条带模式分析的新闻主播检测方法，包括：从经过编辑的新闻视频中截取连续的N帧作为一组，并提取水平时空条带和垂直时空条带；提取水平时空条带和垂直时空条带中所对应的图像特征，得到相应的特征向量；通过聚类方法对特征向量分别聚类，并将同一类中时间连续的水平或垂直时空条带分别合并，作为类中的新元素，得到最终的水平聚类结果和垂直聚类结果；将水平聚类结果中包含有最多元素的类和垂直聚类结果中包含有最多元素的类进行融合，根据融合结果检测新闻主播镜头。本发明的优点是对各类新闻视频主播检测准确率高，通用性强，计算复杂度低。避免了现有方法过于依赖准确的镜头分割和其他模态信息的缺点。



1、一种基于时空条带模式分析的新闻主播检测方法，按以下步骤顺序执行：

步骤 10)、从经过编辑的新闻视频中截取连续的 N 帧作为一组，并从所截取的连续帧组中提取水平时空条带和垂直时空条带；

步骤 20)、分别提取所述水平时空条带和垂直时空条带中所对应的图像特征，得到相应的特征向量；

步骤 30)、通过聚类方法对所述的水平时空条带和垂直时空条带所对应的特征向量分别聚类，并将同一类中时间连续的水平或垂直时空条带分别合并，作为类中的新元素，得到最终的水平聚类结果和垂直聚类结果；

步骤 40)、将所述的水平聚类结果中包含有最多元素的类和所述的垂直聚类结果中包含有最多元素的类进行融合，根据融合结果检测新闻主播镜头。

2、根据权利要求 1 所述的基于时空条带模式分析的新闻主播检测方法，其特征在于，在所述的步骤 10) 中，提取水平时空条带是指：

从所述的连续帧组中提取各连续图像帧水平方向的同一行像素，并将所提取的各行像素拼接成一幅新的图像，所得到的新图像为水平时空条带，所述水平时空条带的长为的一组连续帧组中包含的图像的帧数，宽为一帧图像的长。

3、根据权利要求 1 所述的基于时空条带模式分析的新闻主播检测方法，其特征在于，在所述的步骤 10) 中，提取垂直时空条带是指：

从所述的连续帧组中提取各连续图像帧垂直方向的同一列像素，并将所提取的各列像素拼接成一幅新的图像，所得到的新图像即为垂直时空条带，所述垂直时空条带的长为的一组连续帧组中包含的图像的帧数，宽为一帧图像的宽。

4、根据权利要求 1 所述的基于时空条带模式分析的新闻主播检测方法，其特征在于，在所述的步骤 10) 中，所述 N 的值为 25 的整数倍。

5、根据权利要求 1 所述的基于时空条带模式分析的新闻主播检测方法，其特征在于，在所述的步骤 20) 中，所述的图像特征为颜色特征与纹理特征。

6、根据权利要求 5 所述的基于时空条带模式分析的新闻主播检测方法，其特征在于，提取图像的颜色特征可在颜色空间 RGB 或 HSV 或 HIS 或 YUV 或 Lab 中实现。

7、根据权利要求 6 所述的基于时空条带模式分析的新闻主播检测方法，其特征在于，在所述的颜色空间 HSV 中提取颜色特征时，包括以下步骤：

步骤 21-1)、将图像的 RGB 值转化为色度值、饱和度值和亮度值;

步骤 21-2)、对图像的色度值、饱和度值和亮度值分别进行等级量化;

步骤 21-3)、增大色调的关注度,并将等级量化后的三维向量通过线性组合转换成为一个整数值,每个数值代表一个颜色分段;

步骤 21-4)、将图像中的每个象素都进行量化后,提取图像的颜色特征。

8、根据权利要求 7 所述的基于时空条带模式分析的新闻主播检测方法,其特征在于,在所述的步骤 21-4) 中,提取图像的颜色特征是指:首先将整个图像平均划分为 $4*4$ 的小块,然后将所有小块组合为 5 个大的分块,分别对应上下左右和中间部分,对不同的分块,在提取颜色特征时有以下不同的实现方法:

对上下左右四个板块各提取三维的颜色矩,所述三维的颜色矩包括颜色的一阶矩、二阶矩和三阶矩;

对中间板块提取量化为 32 级的直方图;

对整幅图像还要提取用于描述图像全局颜色特征的颜色三阶矩。

9、根据权利要求 5 所述的基于时空条带模式分析的新闻主播检测方法,其特征在于,提取图像的纹理特征是指:通过对整幅图像提取边缘直方图来描述纹理特征。

10、根据权利要求 1 所述的基于时空条带模式分析的新闻主播检测方法,其特征在于,在所述的步骤 30) 中,所述的聚类方法采用 K-mean 聚类方法。

11、根据权利要求 1 所述的基于时空条带模式分析的新闻主播检测方法,其特征在于,所述的步骤 40) 具体包括以下步骤:

步骤 41)、计算所述的水平聚类结果中包含有最多元素的类和所述的垂直聚类结果中包含有最多元素的类中对应镜头的相似度;

步骤 42)、根据所述的镜头相似度计算结果,判断水平聚类和垂直聚类中的两个镜头的相似度是否大于预先设定的第一阈值,如果是,则两个镜头为类中的对应元素,否则为非对应元素;

步骤 43)、计算所述的水平聚类结果中包含有最多元素的类和所述的垂直聚类结果中包含有最多元素的类的相似度;

步骤 44)、根据步骤 43) 得到的水平聚类结果中包含有最多元素的类和垂直聚类结果中包含有最多元素的类的相似度计算结果,判断两个类的相似度是否大于预先设定的第二阈值,如果是,则将所述的两个类进行融合,作为新闻主播镜头对应的类,如果两个类的相似度小于或等于预先设定的第二阈值,则视频不具有明显的新闻视频结构,无须提取新闻主播镜头。

12、根据权利要求 11 所述的基于时空条带模式分析的新闻主播检测方法，其特征在于，在所述的步骤 41) 中，计算对应镜头的相似度是指：计算对应镜头在时域上相交部分时间长度占总时间长度的比例，若相似度的计算结果小于零，则修正为零。

13、根据权利要求 11 所述的基于时空条带模式分析的新闻主播检测方法，其特征在于，在所述的步骤 43) 中，计算类的相似度是指：通过计算两个类中各元素间的相似度总和来计算两个类的相似度。

14、根据权利要求 11 所述的基于时空条带模式分析的新闻主播检测方法，其特征在于，在所述的步骤 44) 中，两个类进行融合是指：如果一个类中的元素在另一个类中不存在步骤 42) 所述的对应元素，则将该元素作为融合后的类的一个新的元素，如果两个类存在步骤 42) 所述的对应元素，则将对对应元素进行合并，并将两个对应元素的最早开始时间和最晚结束时间作为合并后元素的起始时间。

15、根据权利要求 11 所述的基于时空条带模式分析的新闻主播检测方法，其特征在于，所述第一阈值的值为 0.5，所述第二阈值的值为 0.5。

一种基于时空条带模式分析的新闻主播镜头检测方法

技术领域

本发明涉及视频分析与检测领域，特别涉及一种基于时空条带模式分析的新闻主播检测方法。

背景技术

随着网络的高速发展，人们可以接触大量的图像、视频等可视化数据。根据人们日益增长的对可视化数据的需求，如何对这些海量信息进行有效的分析、管理、查询和检索，成为目前亟待解决的问题。

对于基于内容的视频分析，由于新闻视频特殊的结构特征，研究者将其作为重要的研究对象。一段新闻视频可以被看作主播镜头，新闻故事镜头和可能存在的广告镜头的组合。其中，主播镜头及其后连续的新闻故事单元组成了一个完整的新闻事件。因为主播镜头的内容常常是关于接下来的故事单元的简介，所以新闻主播镜头检测对于新闻视频索引的建立十分重要。

现有新闻主播检测方法主要分为三类：第一类方法是模版匹配法，该方法通过对特定的新闻视频主播建立模版，计算被检测视频帧与模版的相似度来检测新闻主播镜头，该方法鲁棒性差且准确率低；为了增强方法的鲁棒性，研究人员通过第二类融合视觉、听觉等多模态信息的方法建立主播模型来检测新闻主播，尽管该方法在一定程度上提高了鲁棒性，但是存在为每个新闻主播建立模型工作量大和计算复杂度高的缺点；第三类方法不依赖于主播模型的建立，但是它们往往依赖于静态的演播室背景以及准确的镜头分割。总而言之，现有的新闻主播检测方法在鲁棒性和准确性上都存在一定的不足。

目前主要存在两个关键的因素制约新闻主播检测的准确性和鲁棒性：

1、多样的视频节目源造成主播以及演播室背景多样化，这使得目前还没有一种通用的新闻主播检测方法；

2、先进的视频编辑技术使得镜头边界转换形式多样化，如：切变，溶解和淡入淡出等，导致目前并未有通用的方法能够实现准确和自动地提取完整的镜头单元。

发明内容

本发明的目的是克服现有的新闻主播检测方法鲁棒性差、准确率低，计算复杂度高，且通用性不广的缺陷，从而提供一种通用性广、准确率高，计算复杂度低的新闻主播检测方法。

为了实现上述目的，本发明提供了一种基于时空条带模式分析的新闻主播检测方法，按以下步骤顺序执行：

步骤 10)、从经过编辑的新闻视频中截取连续的 N 帧作为一组，并从所截取的连续帧组中提取水平时空条带和垂直时空条带；

步骤 20)、分别提取所述水平时空条带和垂直时空条带中所对应的图像特征，得到相应的特征向量；

步骤 30)、通过聚类方法对所述的水平时空条带和垂直时空条带所对应的特征向量分别聚类，并将同一类中时间连续的水平或垂直时空条带分别合并，作为类中的新元素，得到最终的水平聚类结果和垂直聚类结果；

步骤 40)、将所述的水平聚类结果中包含有最多元素的类和所述的垂直聚类结果中包含有最多元素的类进行融合，根据融合结果检测新闻主播镜头。

上述技术方案中，在所述的步骤 10) 中，提取水平时空条带是指：

从所述的连续帧组中提取各连续图像帧水平方向的同一行像素，并将所提取的各行像素拼接成一幅新的图像，所得到的新图像为水平时空条带，所述水平时空条带的长为一组连续帧组中包含的图像的帧数，宽为一帧图像的长。

上述技术方案中，在所述的步骤 10) 中，提取垂直时空条带是指：

从所述的连续帧组中提取各连续图像帧垂直方向的同一列像素，并将所提取的各列像素拼接成一幅新的图像，所得到的新图像即为垂直时空条带，所述垂直时空条带的长为一组连续帧组中包含的图像的帧数，宽为一帧图像的宽。

上述技术方案中，在所述的步骤 10) 中，所述 N 的值为 25 的整数倍。

上述技术方案中，在所述的步骤 20) 中，所述的图像特征为颜色特征与纹理特征。

在提取图像的颜色特征时，可在颜色空间 RGB (Red, Green, Blue; 红绿蓝三原色) 或 HSV (Hue, Saturation, Value; 色调/饱和度/纯度色彩模型) 或 HIS (Hue, Saturation, Intensity; 色调/饱和度/亮度色彩模型) 或 YUV (Y: 亮度信号; U 和 V: 色差信号) 或 Lab (L: 亮度信号; a 和 b: 色差信号) 中实现。

在所述的颜色空间 HSV 中提取颜色特征时，包括以下步骤：

步骤 21-1)、将图像的 RGB 值转化为色度值、饱和度值和亮度值；

步骤 21-2)、对图像的色度值、饱和度值和亮度值分别进行等级量化；

步骤 21-3)、增大色调的关注度，并将等级量化后的三维向量通过线性组合转换成为一个整数值，每个数值代表一个颜色分段；

步骤 21-4)、将图像中的每个像素都进行量化后，提取图像的颜色特征。

在所述的步骤 21-4) 中，提取图像的颜色特征是指：首先将整个图像平均划分为 4×4 的小块，然后将所有小块组合为 5 个大的分块，分别对应上下左右和中间部分，对不同的分块，在提取颜色特征时有以下不同的实现方法：

对上下左右四个板块各提取三维的颜色矩，所述三维的颜色矩包括颜色的一阶矩、二阶矩和三阶矩；

对中间板块提取量化为 32 级的直方图；

对整幅图像还要提取用于描述图像全局颜色特征的颜色三阶矩。

提取图像的纹理特征是指：通过对整幅图像提取边缘直方图来描述纹理特征。

上述技术方案中，在所述的步骤 30) 中，所述的聚类方法采用 K-mean 聚类方法。

上述技术方案中，所述的步骤 40) 具体包括以下步骤：

步骤 41)、计算所述的水平聚类结果中包含有最多元素的类和所述的垂直聚类结果中包含有最多元素的类中对应镜头的相似度；

步骤 42)、根据所述的镜头相似度计算结果，判断水平聚类和垂直聚类中的两个镜头的相似度是否大于预先设定的第一阈值，如果是，则两个镜头为类中的对应元素，否则为非对应元素；

步骤 43)、计算所述的水平聚类结果中包含有最多元素的类和所述的垂直聚类结果中包含有最多元素的类的相似度；

步骤 44)、根据步骤 43) 得到的水平聚类结果中包含有最多元素的类和垂直聚类结果中包含有最多元素的类的相似度计算结果，判断两个类的相似度是否大于预先设定的第二阈值，如果是，则将所述的两个类进行融合，作为新闻主播镜头对应的类，如果两个类的相似度小于或等于预先设定的第二阈值，则视频不具有明显的新闻视频结构，无须提取新闻主播镜头。

在所述的步骤 41) 中，计算对应镜头的相似度是指：计算对应镜头在时域上相交部分时间长度占总时间长度的比例，若相似度的计算结果小于零，则修正为零。

在所述的步骤 43) 中, 计算类的相似度是指: 通过计算两个类中各元素间的相似度总和来计算两个类的相似度。

在所述的步骤 44) 中, 两个类进行融合是指: 如果一个类中的元素在另一个类中不存在步骤 42) 所述的对应元素, 则将该元素作为融合后的类的一个新的元素, 如果两个类存在步骤 42) 所述的对应元素, 则将对对应元素进行合并, 并将两个对应元素的最早开始时间和最晚结束时间作为合并后元素的起始时间。

所述第一阈值的值为 0.5, 所述第二阈值的值为 0.5。

本发明的优点是对各类新闻视频主播检测准确率高, 通用性强, 计算复杂度低, 避免了现有方法过于依赖准确的镜头分割和其他模态信息的缺点。

附图说明

图 1 为本发明的基于时空条带模式分析的新闻主播检测方法的流程图;

图 2 (a) 为原始视频结构的示意图;

图 2 (b) 是对图 2 (a) 的原始视频结构提取水平时空条带的示意图;

图 2 (c) 是对图 2 (a) 的原始视频结构提取垂直时空条带的示意图;

图 3 是本发明实施例中基于固定分块的图像颜色特征提取方法的示意图。

具体实施方式

下面结合附图和具体实施方式对本发明作进一步详细描述:

如图 1 所示, 本发明的基于时空条带模式分析的新闻主播检测方法包括以下步骤:

步骤 10、从经过编辑的新闻视频中截取连续的 N 帧作为一组, 并从所提取的连续帧组中提取水平时空条带和垂直时空条带。在本步骤中, 在提取水平时空条带时, 如图 2 (a)、图 2 (b) 所示, 提取每组中连续的各帧图像水平方向的同一行像素, 并将所提取的各行像素拼接成一幅新的图像, 所得到的新图像即为水平时空条带。水平时空条带的长为一组中包含图像的帧数, 宽为视频中一帧图像的长。在提取垂直时空条带时, 如图 2 (a)、图 2 (c) 所示, 提取一组中连续的每帧图像垂直方向的同一列像素, 并将所提取的各列像素拼接成一幅新的图像, 所得到的新图像即为垂直时空条带, 垂直时空条带的长为一组中包含的图像的帧数, 宽为视频中一帧图像的宽。

在上述提取水平时空条带和垂直时空条带时，应当提取连续帧组中各个帧的相同行数或相同列数的像素，但所提取的像素行或像素列的位置可以根据需要选择，在本实施例中，可选择帧图像中的中间行或中间列。

在本步骤中，所截取的连续 N 帧的 N 值可根据需要设定，在本实施例中，N 值可取 50。本步骤通过对连续帧的截取，避免了对镜头分割准确检测的依赖。经过编辑的新闻视频经过本步骤后，可得到一系列水平时空条带和垂直时空条带。

步骤 20、提取水平时空条带和垂直时空条带中所对应的图像特征。由于新闻主播镜头中，新闻主播和演播室背景的变化相对新闻故事镜头中的人物和背景变化较小，因此新闻主播镜头对应时空条带的颜色和纹理仅有微小变化，而新闻故事单元对应的时空条带的颜色和纹理变化更为明显，因此采用图像的颜色和纹理特征来表征图像有利于后续的图像聚类操作。从上述的步骤 10 可知，水平时空条带和垂直时空条带都是拼接而成的新图像，对它们做图像特征的提取具体包括以下步骤：

步骤 21、提取图像的颜色特征。在提取图像的颜色特征时，可在不同的颜色空间内进行，如常见的 RGB, HSV, HIS, YUV, Lab 等皆可。在本实施例中，以 HSV (Hue 色度, Saturation 饱和度, Value 亮度) 空间为例，实现对颜色特征的提取。在提取颜色特征时，包括：

步骤 21-1、将图像的 RGB 值转化为色度、饱和度和亮度值；

步骤 21-2、将图像的色度、饱和度和亮度值分别进行等级量化；由于 HSV 空间中，每个像素对应一个三维向量，分别表示所在像素点的色度、饱和度和亮度，但三个值的单位和变化范围都不相同，因此需要分别做等级量化。

步骤 21-3、根据视觉研究的结果增大色调的关注度，并将等级量化后的三维向量通过线性组合转换成为一个 0 到 31 之间的整数值，每个数值代表一个颜色分段。

步骤 21-4、每个像素都进行量化后，提取图像的颜色特征。提取图像颜色特征的具体实现如图 3 所示，将图像平均划分为 4*4 的小块，然后按图 3 所示组合为 5 个大的分块，分别对应上下左右和中间部分（ABCD 分别对应上、左、下、右四个部分，E 对应中间部分，粗线条表示这些大的分块的边界）。对不同的分块，在提取颜色特征时有不同的实现方法：

对上下左右四个板块各提取 3 维的颜色矩（颜色的一阶矩、二阶矩和三阶矩）；

对中间板块提取量化为 32 级的直方图；

对整幅图像还要提取用于描述图像全局颜色特征的颜色三阶矩。

步骤 22、提取图像的纹理特征；在提取图像的纹理特征时，通过对整幅图像提取边缘直方图来描述纹理特征。

上述步骤 21-4 提取颜色特征和步骤 22 提取纹理特征的具体实现是成熟的现有技术，在参考文献《DK Park, YS Jeon, CS Won, and S.-J. Park, Efficient use of local edge histogram descriptor, Proc. of the ACM Workshops on Multimedia, Los Angeles, CA, Nov. 2000》中对纹理特征的提取有详细的记载。

步骤 23、将每个时空条带所有图像特征结合起来形成特征向量，来表征该时空条带的特征。

经过本步骤的上述操作，每个时空条带形成了一个用于表示该条带特征的高维向量，一段新闻视频则根据步骤 10 中所提取的时空条带形成多个特征向量。

步骤 30、通过聚类方法对水平时空条带和垂直时空条带所对应的特征向量分别聚类，并将同一类中时间连续的水平或垂直时空条带分别合并，作为类中的新的元素。

一般而言，新闻视频具有以下结构特征：

- ① 新闻主播镜头经常在新闻视频中周期性出现；
- ② 各新闻主播镜头常具有很高的相似度；
- ③ 新闻主播镜头对应的类应具有最多的元素，因为主播镜头常常具有很高的视觉相似度，而其他镜头仅与同一故事单元中的时域相邻的镜头相似。

根据新闻视频的上述结构特征，本发明制定如下规则：新闻主播镜头是新闻视频中唯一一类在整个节目中具有很多与其具有相似视觉内容的镜头。由上述规则可知，水平和垂直聚类结果中分别含有最多元素的类就是包含新闻主播镜头的类，其中，每个时空条带对应视频中的一个镜头。用 $Cluster_{Max}^H$ 和 $Cluster_{Max}^V$ 分别表示水平和垂直方向聚类结果中包含元素最多的类，它们可以用下式表示：

$$Cluster_{Max}^H = \langle Shot_1^H, Shot_2^H, \dots, Shot_R^H \rangle$$

$$Cluster_{Max}^V = \langle Shot_1^V, Shot_2^V, \dots, Shot_S^V \rangle$$

其中 R 和 S 分别表示水平/垂直聚类结果中包含元素最多的类的元素数。

此外，在新闻故事镜头中存在某些人物长期出现在镜头前且变化的情况，因此为了防止此类情况与新闻主播镜头间的混淆，在本步骤中把那些同一类中时间连续的水平或垂直时空条带分别合并，作为类中的新的元素。

本步骤做聚类操作时可采用 K-mean 聚类方法。

步骤 40、将步骤 30 中所得到的水平聚类结果中包含有最多元素的类和垂直聚类结果中包含有最多元素的类进行融合，根据融合结果检测新闻主播镜头。在以下描述中，用 $Cluster_{Max}^H$ 表示水平聚类结果中包含有最多元素的类，用 $Cluster_{Max}^V$ 表示垂直聚类结果中包含有最多元素的类，本步骤的具体操作如下：

步骤 41、计算 $Cluster_{Max}^H$ 和 $Cluster_{Max}^V$ 中对应镜头的相似度，在计算镜头相似度时，通过计算 $Shot_i^H$ 和 $Shot_j^V$ 在时域上相交部分时间长度占总时间长度的比例来计算两个镜头的相似度。相似度的计算公式如下：

$$Sim < Shot_i^H, Shot_j^V > = \frac{Min(T_{End}^H, T_{End}^V) - Max(T_{Start}^H, T_{Start}^V)}{Max(T_{End}^H, T_{End}^V) - Min(T_{Start}^H, T_{Start}^V)}$$

其中 $T_{Start}^H, T_{End}^H, T_{Start}^V, T_{End}^V$ 分别表示 $Shot_i^H$ 和 $Shot_j^V$ 的开始时间和结束时间，Min 和 Max 分别表示取最小值和最大值操作。如果两个公式中计算出的相似度小于 0，则修正为 0。

步骤 42、根据步骤 41 得到的镜头相似度计算结果，判断两个镜头 $Shot_i^H$ 与 $Shot_j^V$ 的相似度 $Sim < Shot_i^H, Shot_j^V >$ 是否大于预先设定的阈值 Th_1 ，如果是，则这两个镜头称为类中的“对应”元素，否则称为“非对应”元素。

步骤 43、计算 $Cluster_{Max}^H$ 和 $Cluster_{Max}^V$ 的相似度，在计算相似度时，通过计算 $Cluster_{Max}^H$ 和 $Cluster_{Max}^V$ 中各元素间的相似度的总和来计算两个类的相似度。其计算公式如下：

$$Sim < Cluster_{Max}^H, Cluster_{Max}^V > = \frac{1}{Min(R, S)} \sum_{i=1}^R \sum_{j=1}^S Sim < Shot_i^H, Shot_j^V >$$

步骤 44、根据步骤 43 得到的 $Cluster_{Max}^H$ 和 $Cluster_{Max}^V$ 的相似度计算结果，判断 $Cluster_{Max}^H$ 和 $Cluster_{Max}^V$ 的相似度 $Sim < Cluster_{Max}^H, Cluster_{Max}^V >$ 是否大于预先设定的阈值 Th_2 ，如果是，则融合类 $Cluster_{Max}^H$ 和 $Cluster_{Max}^V$ ，作为最终新闻主播对应的类，如果 $Cluster_{Max}^H$ 和 $Cluster_{Max}^V$ 的相似度 $Sim < Cluster_{Max}^H, Cluster_{Max}^V >$ 小于或等于预先设定的阈值 Th_2 ，此时表示该视频不具有明显的新闻视频结构，无须提取新闻主播镜头。

在上述的融合过程中，如果 $Cluster_{Max}^H$ 或 $Cluster_{Max}^V$ 中一个元素在另一类中不存在步骤 42 中所得到的“对应”元素，则将该元素作为最终类的一个新的元素，如

果两个类中存在“对应”元素，则将“对应”元素合并，将两个对应元素的最早的开始时间和最晚的结束时间作为新元素的起始时间。

在步骤 42 中所采用的阈值 Th_1 的值可取 0.5，在步骤 44 中所采用的阈值 Th_2 的值也可取 0.5，但根据实际情况这两个阈值可以做适当的调整。

最后所应说明的是，以上实施例仅用以说明本发明的技术方案而非限制。尽管参照实施例对本发明进行了详细说明，本领域的普通技术人员应当理解，对本发明的技术方案进行修改或者等同替换，都不脱离本发明技术方案的精神和范围，其均应涵盖在本发明的权利要求范围当中。

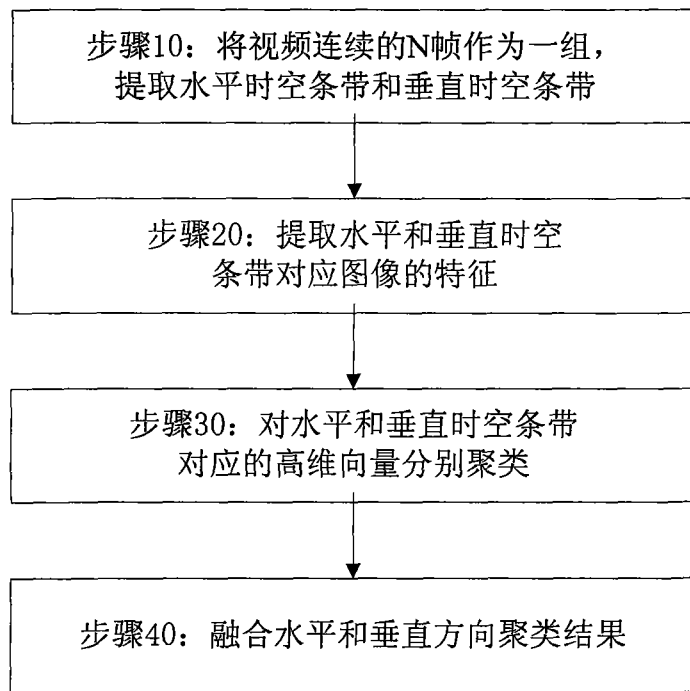


图 1

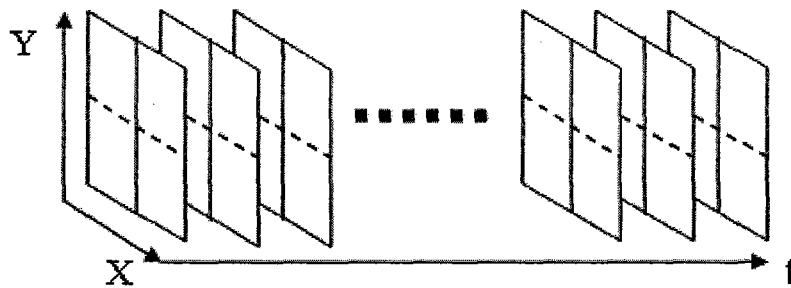


图 2 (a)

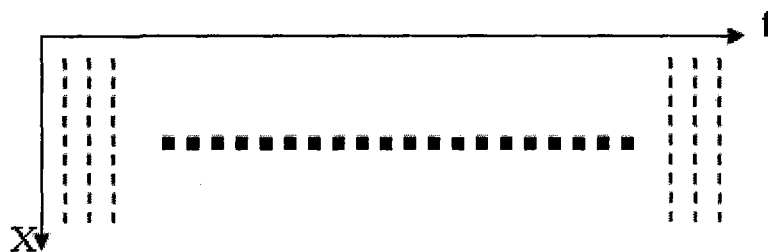


图 2(b)

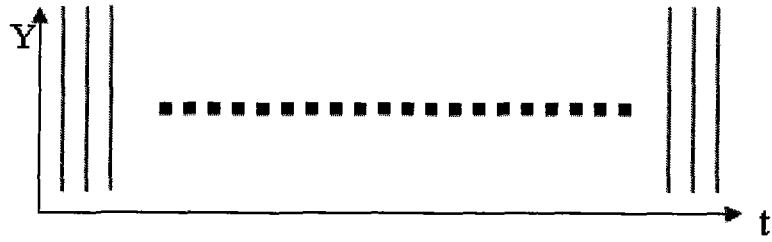


图 2(c)

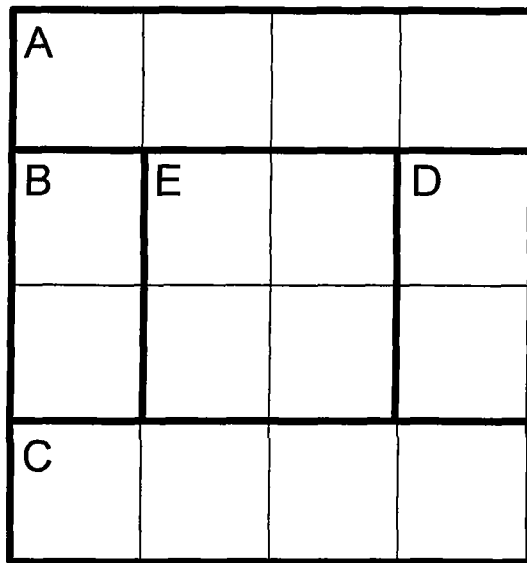


图 3