[19] 中华人民共和国国家知识产权局

[51] Int. Cl[7]
H04N 5/00
H04N 5/93 G11B 27/10
G06F 17/30

[12] 发明专利申请公开说明书

[21] 申请号 03148305.4

[43] 公开日 2003 年 12 月 10 日

[11] 公开号 CN 1461142A

[22] 申请日 2003.6.30 [21] 申请号 03148305.4

[71] 申请人 北京大学计算机科学技术研究所
地址 100871 北京市海淀区北京大学计算机科学技术研究所
共同申请人 北京北大方正技术研究院有限公司
[72] 发明人 彭宇新 杨宗桦 肖建国

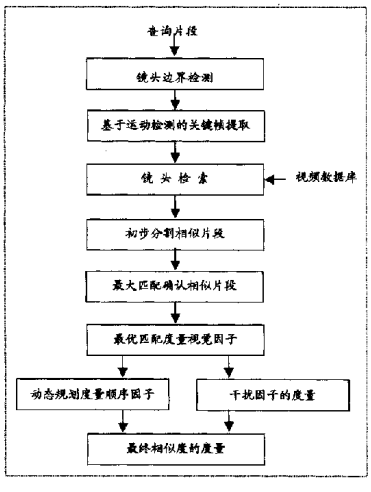
[74] 专利代理机构 北京英赛嘉华知识产权代理有限责任公司
代理人 田 明 王达佐

权利要求书 3 页 说明书 12 页 附图 3 页

[54] 发明名称 一种基于内容的视频片段检索方法

[57] 摘要

本发明属于视频检索技术领域，具体涉及一种基于内容的视频片段检索方法。现有的基于内容的视频片段检索方法往往存在着检索精度不高，检索速度慢的问题。针对现有技术中存在的不足，本发明首次运用图论的最大匹配和最优匹配来解决这个问题。首先，通过考察相似镜头的连续性初步得到一个个相似片段，再运用最大匹配的 Hungarian 算法来确定真正的相似片段。然后，本发明提出用最优匹配的 Kuhn – Munkres 算法和动态规划算法相结合，来解决片段相似度的度量问题。实践结果表明，与现有方法相比，本发明可以取得更高的检索精度和更快的检索速度，同时在相似片段的排列顺序上，更加符合人的心理特征。



1、一种基于内容的视频片段检索方法，包括以下步骤：

（1）首先进行镜头边界检测，把查询片段和视频库中的视频分割为镜头；然后度量查询片段的镜头和视频数据库的镜头的相似度，根据度量结果，检索出视频数据库中与查询片段的镜头相似的所有镜头；

（2）通过考察相似镜头的连续性，初步分割出与查询片段相似的片段；

（3）这些片段包括了真正相似的片段和不相似的片段，此时图论的最大匹配被使用来过滤不相似的片段，而仅仅保留相似的片段到下一步；

（4）对于相似片段，图论的最优匹配计算它们和查询片段的视觉相似度即视觉因子；基于最优匹配的结果，动态规划算法度量两个相似片段时间顺序的相似性即顺序因子；干扰因子也被进一步度量；最终两个片段的相似度表示为上述视觉因子、顺序因子和干扰因子的线性组合。

2、如权利要求1所述的一种基于内容的视频片段检索方法，其特征在于：在进行视频片段检索时，将图论中二分图的理论、算法及结果引入到视频内容的相似度度量上，具体来说，是将图论中最大匹配的 Hungarian 算法和最优匹配的 Kuhn-Munkres 算法用于基于内容的视频片段检索。

3、如权利要求2所述的一种基于内容的视频片段检索方法，其特征在于：步骤（3）中，利用最大匹配的 Hungarian 算法来过滤不相似的片段和确定真正的相似片段：对于二分图 $G_k=\{X,Y_k,E_k\}$，如果 $|M|\geq\lceil n/2\rceil$，则片段 Y_k 与查询片段 X 相似，上述计算式中，$E_k=\{e_{ij}\}$ e_{ij} 表示 x_i 与 y_j 相似，最大匹配 $M\subseteq E_k$，并且 M 中任意两条边都不相邻，n 是查询片段 X 的镜头数。

4、如权利要求2所述的一种基于内容的视频片段检索方法，其特征在于：步骤（4）中，利用最优匹配的 Kuhn-Munkres 算法和动态规划算法具体计算两个片段的相似度：最优匹配的 Kuhn-Munkres 算法计算查询片段 X 和相似片段 Y_k 的视觉因子 $Vision=\frac{\omega}{n}$，式中 ω 为带权二分图 $G_k=\{X,Y_k,E_k\}$ 的最大权，n 是查询片段 X 的镜头数；基于最优匹配的结果，动态规划算法度量两个相似片段的顺序因子 $order=\frac{c[i,j]}{n}$；干扰因子也被进一步度量：$Interference=\frac{2\times|M|}{n+l}$，式中 l 是相似片段 Y_k' 的镜头数目，$|M|$ 表示 $G_k=\{X,Y_k,E_k\}$ 最优匹配的边数，最终两个片段的相似度表示为上述视觉因子、顺序因子和干扰因子的线性组合：$Similarity(X,Y_k')=\omega_1\cdot Vision+\omega_2\cdot Order+\omega_3\cdot Interference$，该式中的 ω_1、ω_2、ω_3 分别表示视觉、顺序、干扰因子的权重。

5、如权利要求1所述的一种基于内容的视频片段检索方法，其特征在于：步骤（2）中，初步分割出视频库 Y 中与查询片段 X 相似的片段：将视频库 Y 中

与查询片段X相似的镜头y_j从小到大排序，然后考察这些y_j的连续性，如果$|y_{j+1}-y_j|>2, j=1,2,...,\lambda-1$，则得到一个可能的相似片段$Y_k=\{y_i,y_{i+1},...,y_j\}$，$i,j\in[1,\lambda]$，上述式子中，$\lambda$是视频库$Y$的长度，以镜头数表示。

6、如权利要求4所述的一种基于内容的视频片段检索方法，其特征在于最优匹配计算视觉因子和确定相似片段的边界的方法如下：

把每对相似镜头的相似值作为权值赋给$G_k=\{X,Y_k,E_k\}$的每条边，这时的G_k就转化为一个带权的二分图，具体计算最优匹配的Kuhn_Munkres算法如下：

（1）给出初始标号$l(x_i)=\max_j \omega_{ij}, l(y_j)=0, i,j=1,2...,t$，$t=\max(n,m)$；

（2）求出边集$E_l=\{(x_i,y_j)\,|\,l(x_i)+l(y_j)=\omega_{ij}\}$、$G_l=(X,Y_k,E_l)$及$G_l$中的一个匹配$M$；

（3）如M已饱和X的所有结点，则M即是G的最优匹配，计算结束，否则进行下一步；

（4）在X中找一M非饱和点x_0，令$A\leftarrow\{x_0\}, B\leftarrow\phi$，$A,B$是两个集合；

（5）若$N_{G_l}(A)=B$，则转第（9）步，否则进行下一步，其中，$N_{G_l}(A)\subseteq Y_k$，是与A中结点邻接的结点集合；

（6）找一结点$y\in N_{G_l}(A)-B$；

（7）若y是M饱和点，则找出y的配对点z，令$A\leftarrow A\cup\{z\}, B\leftarrow B\cup\{y\}$，转第（5）步，否则进行下一步；

（8）存在一条从x_0到y的可增广路P，令$M\leftarrow M\oplus E(P)$，转第（3）步；

（9）按下式计算a值：$a=\min_{\substack{x_i\in A\\ y_j\notin N_{G_l}(A)}}\{l(x_i)+l(y_j)-\omega_{ij}\}$，修改标号：

$$l'(v)=\begin{cases} l(v)-a, 若v\in A \\ l(v)+a, 若v\in B \\ l(v), 其它 \end{cases},$$

根据l'求$E_{l'}$及$G_{l'}$；

（10）$l\leftarrow l', G_l\leftarrow G_{l'}$，转第（6）步；

求出最大权ω和取得ω的匹配M后，视觉因子$Vision=\frac{\omega}{n}$；为了确定Y_k与X相似的片段边界，本发明取X关联M的所有y，从小到大排序为$\{y_\alpha,y_\beta,...,y_\gamma\}, \alpha,\beta,\gamma\in[1,m]$，在这个集合中，$y_\alpha,y_\beta$可能并不连续，即$y_\beta-y_\alpha>1$，根据视频片段连续性的定义，本发明取$y_\alpha$与$y_\gamma$之间的所有镜头构成相似片段$Y_k'=\{y_\alpha,y_{\alpha+1},...,y_\gamma\}$。

7、如权利要求4所述的一种基于内容的视频片段检索方法，其特征在于动态规划算法计算顺序因子的方法如下：

在计算的最优匹配M中，进一步考察Y_k'和X按时间顺序对应的情况，即

找到$Y_k^{'}$按时间顺序和X有边的最长镜头数目，以此来度量顺序因子；这个问题可以归结为最长公共子序列(LCS)问题：给定两个序列$X=\{x_1,x_2,...,x_n\}$和$Y_k^{'}=\{y_\alpha,y_{\alpha+1},...,y_\gamma\}$，要求找出$X$和$Y_k^{'}$的一个最长公共子序列，动态规划算法可以有效解决这个问题，为了计算方便，我们把$\{y_\alpha,y_{\alpha+1},...,y_\gamma\}$表示为$\{y_1,y_2,...,y_l\},l=\gamma-\alpha+1$，用$c[i,j]$记录序列$X$和$Y_k^{'}$的最长公共子序列的长度，建立递归关系如下：

$$c[i,j]=\begin{cases}0 & \text{当 } i=0 \text{ 或 } j=0\\ c[i-1,j-1]+1 & \text{当 } i,j>0 \text{ 且 } (x_i,y_j)\in M\\ \max(c[i,j-1],c[i-1,j]) & \text{当 } i,j>0 \text{ 且 } (x_i,y_j)\notin M\end{cases}$$

顺序因子$order=\frac{c[i,j]}{n}$。

一种基于内容的视频片段检索方法

技术领域

本发明属于视频检索技术领域，具体涉及一种基于内容的视频片段检索方法。

背景技术

随着电视台视频节目的积累，网上数字视频的增加，以及数字图书馆，视频点播，远程教学等大量的多媒体应用，如何在海量视频中快速检索出所需要的资料显得至关重要。传统的基于关键词描述的视频检索因为描述能力有限，主观性强，手工标注，直观性差等原因，已经不能满足海量视频检索的需求. 因此，从 90 年代开始，基于内容的视频检索技术成为研究的热点问题.

基于内容的视频片段检索是基于内容的视频检索的主要方式，它是指给定一个查询片段，从视频库里找到所有与它相似的片段。基于内容的视频片段检索需要解决两个问题和同时进行两种类型片段的检索。两个问题是：1、从视频库里自动分割出与查询片段相似的多个片段；2、按照相似度从高到低排列这些相似片段。两种类型的检索包括：1、精确检索：要检索的片段与查询片段基本一样，具有同样的镜头和帧序列；2、相似性检索：有这样两种情况，一种是对原视频进行了各种编辑，如插入/删除帧（慢镜头/快镜头）、插入/删除镜头、交换帧/镜头顺序等。另一种是不同拍摄的同类节目，如不同的足球比赛等。一个好的片段检索算法，应该能够解决上述两个问题，同时在合理的时间内进行两种类型片段的检索。

已有的片段检索方法可以分为两类：一、如文献“A Framework for Measuring Video Similarity and Its Application to Video Query by Example”[Y. P. Tan, S. R. Kulkarni, and P. J. Ramadge, IEEE International Conference on Image Processing , Vol. 2, pp. 106-110, 1999]所述，把视频片段分为片段 - 帧两层考虑，片段的相似性利用组成它的帧的相似性来直接度量。这类方法的缺点在于限制相似的片段必须遵守同样的时间顺序，而实际的视频节目并不遵守这种约束，因为后期编辑的结果使得相似的片段完全可能具有不同的镜头顺序，如同一个广告的不同编辑，同时这种基于每帧的比较，也使得检索速度比较慢。二、与本发明最为接近的现有技术是 2001 年在 IEEE International Conference on Multimedia and Expo 发表的文献“A Match and Tiling Approach to

Content-based Video Retrieval" (作者是L. Chen, and T. S. Chua, 页码 417-420), 该对比文献公开了一类片段检索方法, 该方法把视频片段分为片段 - 镜头 - 帧三层考虑,它包括这样几个步骤: (1) 先使用MRA(Temporal Multi-Resolution Analysis)方法检测镜头边界, 然后对每个镜头的每一帧, 进行颜色编码和纹理编码。颜色编码采用 Y 分量的均值μ和方差σ编码, 纹理采用分形维特征(Fractal Dimension, FD)编码; (2) 假设两个镜头内部的相似帧, 按照时间顺序对应相似, 因此计算两个镜头相似帧的最长序列, 最终两个镜头的相似度, 表示为上述 3 个特征的线性组合, 确定相似阈值σ_L, 判断两个镜头是否相似; (3) 在此基础上, 使用滑动窗口(Sliding Window)的办法, 最终找到与查询片段相似的片段。这个方法能够同时进行精确检索和相似性检索, 但它的问题在于: (1)只考虑了两个片段相似镜头的数量, 而没有考虑多对多的镜头相似(粒度)对总体相似度的影响, 因此, 即使片段 *Y*的所有镜头仅仅和片段 *X*的一个镜头相似, *Y*也会被认为与 *X*相似; (2) 提出的假设并不成立, 即两个镜头内部的相似帧, 未必按照时间顺序对应相似; (3)镜头的相似性是根据两个镜头相似的最长帧序列来判断,这种基于每帧的比较,片段的检索速度比较慢。

发明内容

针对现有的视频片段检索方法所存在的缺陷, 本发明的目的是提出一种基于内容的视频片段检索方法, 该方法能在现有技术的基础上大大提高基于内容的视频片段检索的检索精度和检索速度, 从而更加充分地发挥视频片段检索技术在当今网络信息社会中的巨大作用。本发明的另外一个目的是在提高检索精度和检索速度的同时, 在相似片段的排列顺序上, 更加符合人的心理特征。

本发明的目的是这样实现的: 一种基于内容的视频片段检索方法, 包括以下步骤:

(1) 首先使用时空切片算法(spatio-temporal slice)进行镜头边界检测, 把查询片段和视频库中的视频分割为镜头; 然后检测镜头内的相机运动信息, 抽取或构造关键帧来表示镜头内容; 镜头的相似性度量是基于查询片段镜头的关键帧和视频数据库镜头的关键帧比较的结果, 根据镜头检索结果, 检索出视频数据库中与查询片段的镜头相似的所有镜头;

(2) 通过考察相似镜头的连续性, 初步分割出与查询片段相似的片段;

(3)这些片段包括了真正相似的片段和不相似的片段, 此时最大匹配的Hungarian 算法被使用来过滤不相似的片段, 而仅仅保留相似的片段到下一步;

(4)对于相似片段, 图论的最优匹配计算它们和查询片段的视觉相似度

即视觉因子；基于最优匹配的结果，动态规划算法度量两个相似片段时间顺序的相似性即顺序因子；干扰因子也被进一步度量；最终两个片段的相似度表示为上述视觉因子、顺序因子和干扰因子的线性组合。

需要说明的是，因为最优匹配是在一对一（粒度）的前提下，计算得到视觉因子，而顺序因子和干扰因子的计算也是基于最优匹配的结果，所以最终的相似度度量，实际上已经包含了粒度因子的度量。

为了更好地实现本发明的目的，在进行视频片段检索时，将图论中二分图的理论，算法及结果引入到视频内容的相似度度量上，具体来说，是将图论中最大匹配的Hungarian算法和最优匹配的Kuhn-Munkres算法用于基于内容的视频片段检索。

具体来说，在进行视频片段检索时，初步分割出视频库Y中与查询片段X相似的片段：将视频库Y中与查询片段X相似的镜头y_j从小到大排序，然后考察这些y_j的连续性，如果$|y_{j+1}-y_j|>2, j=1,2,...,\lambda-1$，则得到一个可能的相似片段$Y_k=\{y_i,y_{i+1},...,y_j\}$，$i,j\in[1,\lambda]$，上述式子中，$\lambda$是视频库$Y$的长度，以镜头数表示。

再具体来说，在进行视频片段检索时，利用最大匹配的Hungarian算法来过滤不相似的片段和确定真正的相似片段：对于二分图$G_k=\{X,Y_k,E_k\}$，如果$|M|\geq\lceil n/2\rceil$，则片段Y_k与查询片段X相似，上述计算式中，$E_k=\{e_{ij}\}$，e_{ij}表示x_i与y_j相似，最大匹配$M\subseteq E_k$，并且M中任意两条边都不相邻，n是查询片段X的镜头数。

更进一步，在进行视频片段检索时，利用最优匹配的Kuhn-Munkres算法和动态规划算法具体计算两个片段的相似度：最优匹配的Kuhn-Munkres算法计算查询片段X和相似片段Y_k的视觉因子$Vision=\frac{\omega}{n}$，式中ω为带权二分图$G_k=\{X,Y_k,E_k\}$的最大权，n是查询片段X的镜头数；基于最优匹配的结果，动态规划算法度量两个相似片段的顺序因子$order=\frac{c[i,j]}{n}$；干扰因子也被进一步度量：$Interference=\frac{2\times|M|}{n+l}$，式中$l$是相似片段$Y_k'$的镜头数目，$|M|$表示$G_k=\{X,Y_k,E_k\}$最优匹配的边数，最终两个片段的相似度表示为上述视觉因子、顺序因子和干扰因子的线性组合：$Similarity(X,Y_k')=\omega_1\cdot Vision+\omega_2\cdot Order+\omega_3\cdot Interference$，该式中的$\omega_1$、$\omega_2$、$\omega_3$分别表示视觉、顺序、干扰因子的权重。

再具体来说，在进行视频片段检索时，最优匹配计算视觉因子和确定相似片段边界的方法如下：把每对相似镜头的相似值作为权值赋给$G_k=\{X,Y_k,E_k\}$的每条边，这时的G_k就转化为一个带权的二分图，具体计算最优匹配的Kuhn_Munkres算法如下：

（1）给出初始标号$l(x_i)=\max_j\omega_{ij}, l(y_j)=0, i,j=1,2...,t$，$t=\max(n,m)$；

（2）求出边集$E_l=\{(x_i,y_j)\,|\,l(x_i)+l(y_j)=\omega_{ij}\}$、$G_l=(X,Y_k,E_l)$及$G_l$中的一个匹

配M；

（3）如M已饱和X的所有结点，则M即是G的最优匹配，计算结束，否则进行下一步；

（4）在X中找一M非饱和点x_0，令$A \leftarrow \{x_0\}, B \leftarrow \phi$，$A,B$是两个集合；

（5）若$N_{G_l}(A)=B$，则转第（9）步，否则进行下一步，其中，$N_{G_l}(A) \subseteq Y_k$，是与A中结点邻接的结点集合；

（6）找一结点$y \in N_{G_l}(A)-B$；

（7）若y是M饱和点，则找出y的配对点z，令$A \leftarrow A \cup \{z\}, B \leftarrow B \cup \{y\}$，转第（5）步，否则进行下一步；

（8）存在一条从x_0到y的可增广路P，令$M \leftarrow M \oplus E(P)$，转第（3）步；

（9）按下式计算a值：$a=\min_{\substack{x_i \in A \\ y_j \notin N_{G_l}(A)}} \{l(x_i)+l(y_j)-\omega_{ij}\}$，修改标号：

$$l'(v)=\begin{cases} l(v)-a, \text{若} v \in A \\ l(v)+a, \text{若} v \in B \\ l(v), \text{其它} \end{cases},$$

根据l'求$E_{l'}$及$G_{l'}$；

（10）$l \leftarrow l', G_l \leftarrow G_{l'}$，转第（6）步。

求出最大权ω和取得ω的匹配M后，视觉因子$Vision=\frac{\omega}{n}$；为了确定Y_k与X相似的片段边界，本发明取X关联M的所有y，从小到大排序为$\{y_\alpha, y_\beta, ..., y_\gamma\}, \alpha, \beta, \gamma \in [1,m]$，在这个集合中，$y_\alpha, y_\beta$可能并不连续，即$y_\beta - y_\alpha > 1$，根据视频片段连续性的定义，本发明取$y_\alpha$与$y_\gamma$之间的所有镜头构成相似片段$Y_k' = \{y_\alpha, y_{\alpha+1}, ..., y_\gamma\}$。

为了更好地实施本发明，动态规划算法计算顺序因子的方法可以是：在计算的最优匹配M中，进一步考察Y_k'和X按时间顺序对应的情况，即找到Y_k'按时间顺序和X有边的最长镜头数目，以此来度量顺序因子。这个问题可以归结为最长公共子序列(LCS)问题：给定两个序列$X=\{x_1, x_2, ..., x_n\}$和$Y_k' = \{y_\alpha, y_{\alpha+1}, ..., y_\gamma\}$，要求找出$X$和$Y_k'$的一个最长公共子序列，动态规划算法可以有效解决这个问题。为了计算方便，我们把$\{y_\alpha, y_{\alpha+1}, ..., y_\gamma\}$表示为$\{y_1, y_2, ..., y_l\}, l=\gamma-\alpha+1$，用$c[i,j]$记录序列$X$和$Y_k'$的最长公共子序列的长度，建立递归关系如下：

$$c[i,j]=\begin{cases} 0 & \text{当} i=0 \text{或} j=0 \\ c[i-1,j-1]+1 & \text{当} i,j>0 \text{且} (x_i,y_j) \in M \\ \max(c[i,j-1], c[i-1,j]) & \text{当} i,j>0 \text{且} (x_i,y_j) \notin M \end{cases}$$

顺序因子 $order = \frac{c[i,j]}{n}$。

本发明的效果在于：采用本发明所述的视频片段检索方法，可以取得更高的检索精度和更快的检索速度，本发明的另一个效果在于本发明同时在相似片段的排列顺序上，更加符合人的心理特征。

本发明之所以具有如此显著的技术效果，其原因在于：

一、如前面技术内容所述，为了分割出相似片段，本发明把检索过程分为镜头检索和片段检索两个阶段：在镜头检索阶段，考虑了视频中的时间信息，把一个镜头内部随时间变化的内容，分解为几个内容一致的子镜头（sub-shots），这种基于子镜头的比较全面地反映了两个镜头是否相似，它不仅避免了现有方法对每个镜头仅仅采用一个关键帧比较的不足，也避免了现有方法逐帧比较造成的检索速度慢的问题；在片段检索阶段，通过考察相似镜头的连续性初步得到一个个相似片段，再运用最大匹配的 Hungarian 算法来确定真正的相似片段。为了排列相似片段，本发明考虑了片段相似度度量的视觉、粒度、时间顺序和干扰因子，提出用最优匹配的 Kuhn_Munkres 算法和动态规划算法相结合来度量这些因子的影响。本发明首次运用图论的匹配理论来解决视频检索问题，这是因为匹配的思想要求相似镜头必须一一对应（粒度），在这个条件下，求出的最大匹配和最优匹配可以客观全面地反映两个片段相似的镜头数量和两个片段视觉相似的程度，从而避免了现有方法中镜头计算的粒度问题。实验结果表明，与具有同样功能的现有方法相比，无论是检索的准确性，还是检索速度，本发明都取得了出色的效果。

二、视频片段的相似度度量，除了视觉信息以外，还依赖于组成片段的镜头之间的内部关系，为了达到本发明所述的显著技术效果，本发明在具体检索时，考虑了下列 4 个因子：

（1）视觉因子：是决定两个片段是否相似的最重要因素，主要通过组成片段的镜头的相似性来度量；

（2）粒度因子：一个片段里的某个镜头可能会相似于另一个片段里的多个镜头。因此，在两个片段的相似镜头对应图中，会出现一对多、多对一、多对多的情况。需要方法来度量不同镜头对应关系的相似性。例如，两个多对一关系的片段应该被给予更低的相似值；

（3）顺序因子：两个视觉上相似的片段，不能因为不同的镜头顺序而被认为不相似。但是，相比较视觉相似而时间顺序不同的两个片段，视觉和时间顺序都相似的两个片段应该被赋予更高的相似值；

（4）干扰因子：两个相似片段，它们中的一些镜头可能不能找到对应的相似镜头，这些镜头的存在体现了对应的不连续性，对两个片段最终的相似性会产生影响。

三、提出了基于内容的视频片段的检索策略：先找出视觉上与查询片段相似的所有片段；对于相似片段，再计算它们和查询片段的具体相似度，因为视觉是度量两个片段是否相似的最重要因素，这种检索策略的优点在于：视觉上

相似的片段不会因为其它因子的影响而漏掉，同时可以加快检索速度，因为不相似的片段就不用计算它们的具体相似度。

附图说明

图 1 是本发明的总体框架，是本发明中各步方法的流程示意图；

图 2 是两个不相似片段的二分图；

图 3 是两个不相似片段的二分图；

图 4 是两个相似片段的二分图；

图 5 是对图 3 使用求最大匹配的 Hungarian 算法的结果；

图 6 是对图 4 使用求最大匹配的 Hungarian 算法的结果；

图 7 是本发明对一个视频片段的检索结果。

具体实施方式

下面结合附图对本发明作进一步详细的描述。

图 1 列出了本发明各步方法的流程示意图，包括以下步骤：

1、镜头检索

首先使用时空切片算法（spatio-temporal slice）进行镜头边界检测，把查询片段 X 和视频库 Y 中的视频分割为镜头，关于时空切片算法的详细描述可以参考文献“Video Partitioning by Temporal Slice Coherency” [C. W. Ngo, T. C. Pong, and R. T. Chin, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 11, No. 8, pp. 941-953, August, 2001]；然后根据文献“Motion-based Video Representation for Scene Change Detection” [C. W. Ngo, T. C. Pong, and H. J. Zhang, International Journal of Computer Vision, Vol. 50, No. 2, pp.127-143, Nov 2002]中的方法，检测镜头内的相机运动信息，抽取或构造关键帧来表示镜头内容；两个镜头的相似值 $Similarity(x_i, y_j)$ 是根据两个镜头的关键帧计算的结果（其中 x_i, y_j 表示两个镜头）；接着，本发明设定阈值 $T=0.5$，当 $Similarity(x_i, y_j) > T$，就认为两个镜头 x_i 和 y_j 相似，根据这个公式，检索出视频数据库 Y 中与查询片段 X 的镜头 x_i 相似的所有镜头 y_j；

2、初步分割相似片段

对视频库 Y 而言，与查询片段 X 相似的镜头是少数，大量的镜头并不相似。根据片段由连续镜头组成的定义，本发明首先将 Y 中与 X 相似的镜头 y_j 从小到大排序，然后考察这些 y_j 的连续性，如果 $|y_{j+1} - y_j| > 2, j = 1,2,...,\lambda - 1$，$\lambda$ 是视频库 Y 的长度（以镜头数表示），则得到一个可能的相似片段 $Y_k = \{y_i, y_{i+1}, ..., y_j\}$，$i, j \in [1, \lambda]$。我们取 $|y_{j+1} - y_j| > 2$，是考虑算法的鲁棒性，因为：（1）后期编辑会插入无关镜头，如同一个广告的编辑，长广告会在短广告的基础上插入少量不相似的镜头；（2）如果开始一个新片段，它们之间会有一段时间的间隔，这种间隔一般大于 2 个镜头。

3、最大匹配确认相似片段

假设查询片段$X=\{x_1,x_2,...,x_n\}$，每个可能的相似片段$Y_k=\{y_1,y_2,...,y_m\}$，其中x_i,y_j表示镜头，那么，X与Y_k的相似镜头对应图，可以表示为图论中的二分图$G_k=\{X,Y_k,E_k\}$，其中，顶点集$V_k=X\cup Y_k$，边集$E_k=\{e_{ij}\}$，e_{ij}表示x_i与y_j相似。

经过第2步判断的可能相似片段，包含了不相似片段和真正的相似片段。通过大量的实验观察，可以归纳为图2、图3和图4三种典型情况，其中图2和图3是不相似片段的二分图，图4是相似片段的二分图。由于视频片段是由表示同一个语义的连续镜头组成，因此一个视频片段的内部镜头本身就会相似，我们称这个性质为视频片段的自相似性，由于这种自相似性的存在，X和Y_k的二分图会出现普遍的一对多、多对一、多对多的情况，如图2、3、4所示。判断两个片段是否相似，可以从它们相似镜头的数量来判断，经过第2步的判断，我们知道，基本每个y_j在X中都能找到相似镜头x_i，但因为多对多相似的存在，未必每个x_i在Y_k中都能找到相似镜头y_j。因此，我们考察x_i的相似情况，因为Y_k的长度可能会小于X的长度，考虑算法的鲁棒性，如果X中有一半镜头在Y_k中能找到相似镜头，我们就认为Y_k和X相似的镜头足够多，因此Y_k是X的相似片段，这个方法可以有效辨别图2的情况。但在图3和图4，查询片段$X=\{x_1,x_2,...,x_8\}$都有6个镜头找到相似镜头，如果用上述方法，它们都被判断为相似片段，但图3却是不相似片段的典型情况。

因此，我们进一步观察在Y_k和X一一对应而不是重复对应的情况下，它们的相似情况。对图3、4使用求最大匹配的Hungarian算法，得到图5、6，如果$|M|\geq\lceil n/2\rceil$，我们就认为Y_k与X相似的镜头数足够多，因此它是真正的相似片段，该式中最大匹配$M\subseteq E_k$，并且M中任意两条边都不相邻，n是查询片段X的镜头数。这样从图5、6，我们可以清楚地区分出不相似片段和相似片段。具体的Hungarian算法如下：

（1）任给出图G的一个初始匹配M；

（2）若M已饱和X的所有结点，则M即是最大匹配，计算结束，否则进行下一步；

（3）找X中任一M非饱和点x_0，令

$$A\leftarrow\{x_0\},B\leftarrow\phi,\ A,B\text{是两个集合；}$$

（4）如$N(A)=B$，将x_0作为饱和点（或称为伪饱和点）转第（2）步，否则进行下一步（$N(A)\subseteq Y$，是与A中结点邻接的结点集合）；

（5）找一结点$y\in N(A)-B$；

（6）如y是M饱和点，则找出y的配对点z，令

$$A\leftarrow A\cup\{z\},B\leftarrow B\cup\{y\}$$

转第（4）步，否则进行下一步；

（7）存在一条从x_0到y的可增广道路P，令

$M \leftarrow M \oplus P$（M与P进行环和）

转第（2）步。

4、视频片段的相似度模型

经过第3步的计算，我们已经得到与查询片段视觉上相似的多个片段，接下来考虑按照相似度从高到低排列它们。我们考虑了片段相似度度量的下列因子：

（1）视觉因子：是决定两个片段是否相似的最重要因素，主要通过组成片段的镜头的相似性来度量；

（2）粒度因子：一个片段里的某个镜头可能会相似于另一个片段里的多个镜头。因此，在两个片段的相似镜头对应图中，会出现一对多、多对一、多对多的情况。需要方法来度量不同镜头对应关系的相似性。例如，两个多对一关系的片段应该被给予更低的相似值；

（3）顺序因子：两个视觉上相似的片段，不能因为不同的镜头顺序而被认为不相似。但是，相比较视觉相似而时间顺序不同的两个片段，视觉和时间顺序都相似的两个片段应该被赋予更高的相似值；

（4）干扰因子：两个相似片段，它们中的一些镜头可能不能找到对应的相似镜头，这些镜头的存在体现了对应的不连续性，对两个片段最终的相似性会产生影响。

本发明是基于图论的最优匹配来表示和建模上述的相似度模型，这样做的一个显著优点是，本发明的有效性能够通过最优匹配来验证。另外，因为视觉是相似片段最重要的判断标准，我们不是像现有方法那样采用上述因子的线形组合来判断两个片段是否相似，而是先利用最大匹配得到视觉上的相似片段后，再基于最优匹配来表示和建模相似度模型，这样视觉上相似的片段不会因为其它因子的影响而漏掉，另外，因为最大匹配的计算复杂度低于最优匹配，这样做也可以加快检索的速度。最优匹配和最大匹配一样，都是在粒度的前提下进行计算，下面我们具体计算其它的3个因子：

4.1 最优匹配计算视觉因子

我们把每对相似镜头的相似值作为权值赋给$G_k = \{X, Y_k, E_k\}$的每条边，这时的G_k就转化为一个带权的二分图，具体计算最优匹配的Kuhn-Munkres算法如下：

（1）给出初始标号$l(x_i) = \max_j \omega_{ij}, l(y_j) = 0, i, j = 1,2...,t$，$t = \max(n,m)$；

（2）求出边集$E_l = \{(x_i, y_j) \mid l(x_i) + l(y_j) = \omega_{ij}\}$、$G_l = (X, Y_k, E_l)$及$G_l$中的一个匹配$M$；

（3）如M已饱和X的所有结点，则M即是G的最优匹配，计算结束，否则进行下一步；

（4）在X中找一M非饱和点x_0，令$A \leftarrow \{x_0\}, B \leftarrow \phi$，$A, B$是两个集合；

（5）若 $N_{G_l}(A)=B$，则转第（9）步，否则进行下一步，其中，$N_{G_l}(A)\subseteq Y_k$，是与 A 中结点邻接的结点集合；

（6）找一结点 $y\in N_{G_l}(A)-B$；

（7）若 y 是 M 饱和点，则找出 y 的配对点 z，令 $A\leftarrow A\cup\{z\},B\leftarrow B\cup\{y\}$，转第（5）步，否则进行下一步；

（8）存在一条从 x_0 到 y 的可增广路 P，令 $M\leftarrow M\oplus E(P)$，转第（3）步；

（9）按下式计算 a 值：$a=\min_{\substack{x_i\in A\\ y_j\notin N_{G_l}(A)}}\{l(x_i)+l(y_j)-\omega_{ij}\}$，修改标号：

$$l'(v)=\begin{cases}l(v)-a,\text{若}v\in A\\ l(v)+a,\text{若}v\in B\\ l(v),\text{其它}\end{cases},$$

根据 l' 求 $E_{l'}$ 及 $G_{l'}$；

（10）$l\leftarrow l',G_l\leftarrow G_{l'}$，转第（6）步。

求出最大权 ω 和取得 ω 的匹配 M 后，本发明定义视觉因子 $Vision=\frac{\omega}{n}$。为了确定 Y_k 与 X 相似的片段边界，本发明取 X 关联 M 的所有 y，从小到大排序为 $\{y_\alpha,y_\beta,...,y_\gamma\},\alpha,\beta,\gamma\in[1,m]$，在这个集合中，$y_\alpha,y_\beta$ 可能并不连续，即 $y_\beta-y_\alpha>1$，根据视频片段连续性的定义，本发明取 y_α 与 y_γ 之间的所有镜头构成相似片段 $Y_k'=\{y_\alpha,y_{\alpha+1},...,y_\gamma\}$。

4.2 动态规划算法计算顺序因子

在 4.1 计算的最优匹配 M 中，我们进一步考察 Y_k' 和 X 按时间顺序对应的情况，即找到 Y_k' 按时间顺序和 X 有边的最长镜头数目，以此来度量顺序因子。这个问题可以归结为最长公共子序列(LCS)问题：给定两个序列 $X=\{x_1,x_2,...,x_n\}$ 和 $Y_k'=\{y_\alpha,y_{\alpha+1},...,y_\gamma\}$，要求找出 X 和 Y_k' 的一个最长公共子序列，动态规划算法可以有效解决这个问题。为了计算方便，我们把 $\{y_\alpha,y_{\alpha+1},...,y_\gamma\}$ 表示为 $\{y_1,y_2,...,y_l\},l=\gamma-\alpha+1$，用 $c[i,j]$ 记录序列 X 和 Y_k' 的最长公共子序列的长度，建立递归关系如下：

$$c[i,j]=\begin{cases}0 & \text{当}i=0\text{或}j=0\\ c[i-1,j-1]+1 & \text{当}i,j>0\text{且}(x_i,y_j)\in M\\ \max(c[i,j-1],c[i-1,j]) & \text{当}i,j>0\text{且}(x_i,y_j)\notin M\end{cases}$$

本发明定义顺序因子 $order=\frac{c[i,j]}{n}$。

4.3 计算干扰因子

在最优匹配 M 中，X 和 Y_k' 会有少量镜头没有边关联，这说明这些镜头不能找到对应的相似镜头，它们的存在体现了对应的不连续性，本发明定义干扰因

子 $Interference=\frac{2\times|M|}{n+l}$，$n$是查询片段$X$的镜头数目，$l$是相似片段$Y_k'$的镜头数目。这个等式表明两个相似片段$X$和$Y_k'$的所有镜头中，能找到对应相似镜头的镜头比例。

4.4 计算总的相似度

根据前面的分析，本发明用下列公式计算查询片段X和它的相似片段Y_k'的相似度：$Similarity(X,Y_k')=\omega_1\cdot Vision+\omega_2\cdot Order+\omega_3\cdot Interference$

其中，ω_1、ω_2、ω_3表明了人们对视觉、顺序、干扰因子的重视程度，不同的用户可以根据自己对这3个判断标准的喜好程度来调整它们。在本发明，分别取ω_1=0.4，ω_2=0.3，ω_3=0.3，实验结果表明，这种取法能够符合人们的相似性判断标准.

下面用实验结果来说明本发明在视频片段检索中的优异表现。实验数据是从电视录制的几天节目，这个视频数据库非常具有挑战性，总共有3小时11分钟，4714个镜头，286936帧图像，包括了广告、新闻、体育、电影各种类型的节目，这里面有重复的相同视频片段，如新闻的片头、广告等；也有很多重复的相似视频片段，如体育节目中的不同网球比赛、不同时间长度和编辑的相同广告等。为了验证本发明的有效性，我们使用了现有方法作为实验对比，主要有这样两个原因：1、现有方法是目前所给出的实验数据最好的方法，也是最新的一种方法；2、与本发明功能一致，能够在视频库里自动分割出相似片段，然后按相似度从高到低排列这些相似片段。在视频片段检索中，除了检索的准确性以外，检索速度也是非常重要的一个指标，基于这种考虑，我们也比较了两种方法的检索速度，使用的测试机器是PIII Dual CPU 1G Hz，内存256M。

图7是实验程序的用户界面：上面一行是查询的某条广告，显示的是它的关键帧，下面是检索的结果，按照相似度递减的顺序先后排列。检索出的第一行即是查询的片段，它的相似度当然是最高的，其余的片段按照相似度递减的顺序先后排列。可以看到，排列的相似片段体现了第4步中不同因子的作用，如前3个片段和查询片段在时间顺序上更为相似。具体的实验结果分别在表1和表2给出。

表1 视频片段精确检索的实验结果

查询片段	帧数	本发明			现有的方法		
		查准率	查全率	速度(秒)	查准率	查全率	速度(秒)
1、新闻的片头	832	100%	100%	108	75%	100%	230
2、足球新闻	715	100%	100%	74	100%	100%	196

3、汇源广告	367	100%	100%	167	33.3%	100%	97
4、光明广告	374	100%	100%	89	100%	100%	101
5、福临门广告	432	100%	100%	99	100%	100%	116
平 均	544	100%	100%	107	81.7%	100%	148

从表 1 可以看到，本发明和现有方法都取得了 100%的查全率（recall），但在查准率上（precision），本发明优于现有方法，主要原因在于现有方法仅仅计算两个片段相似镜头的数量，而本发明考虑了相似镜头的对应关系。在检索速度上，本发明快于现有方法，根据我们的实验，总的检索时间基本上是等于相似镜头判断的时间，现有方法采用按时间顺序逐帧比较的办法，而本发明只需比较每个镜头的关键帧，因此本发明的检索速度大大快于现有方法。

表 2 视频片段相似性检索的实验结果

查 询 片 段	帧 数	本 发 明			现 有 的 方 法		
		查准率	查全率	速度(秒)	查准率	查全率	速度(秒)
1、网球比赛	507	100%	50%	49	100%	50%	140
2、医生抢救病人	1806	60%	85.7%	93	50%	50%	507
3、TCL 广告	374	100%	100%	116	85.7%	100%	100
4、脑白金广告	374	100%	100%	129	100%	100%	100
5、厦新广告	374	100%	100%	103	100%	50%	99
平 均	687	92%	87.1%	98	87.1%	70%	189

在表 2，无论是查全率，还是查准率，本发明都优于现有方法，查询片段 1 和 2 是两个难度很大的查询，在我们的视频库中，网球比赛共出现 4 次，本发明漏掉了其中两个，原因是我们使用了蓝色网球场查询，而漏掉的一个的网球场是绿色，另外一个主要是选手和观众镜头，反映蓝色球场的镜头很少，现有方法也同样漏掉了这两个片段。与查询片段 1 类似，查询片段 2 也是一个语义很强而颜色特征很难利用的片段，综合整个片段反映这个语义的基本颜色特征，本发明也取得了不错的检索效果。在检索速度上，本发明同样快于现有方法，查询片段越长，本发明的优势越明显，例如在查询片段 2，本发明的速度比现有方法快了 5 倍多。此外，如图 7 所示，相比较现有方法而言，本发明的显著优势还表现在根据相似度从大到小排列相似片段上，因为除了视觉特征，本发明还考虑了相似片段的不同因子，而现有方法的相似度仅仅取决于相似镜头的数量，通过对几个人的测试结果表明，本发明在相似片段的排序上，更加符合人的视觉特征和心理特征。

通过采集 3 小时 11 分钟的视频节目，并和目前实验效果最好和最新的现有方法进行实验对比，结果表明，采用本发明所述的视频片段检索方法，可

以取得更高的检索精度和更快的检索速度，同时在相似片段的排列顺序上，更加符合人的心理特征。除了表 1 和表 2 列出的 6 个广告查询外，我们又查询了十几个不同编辑的广告，本发明都取得了 100%的查准率和查全率。

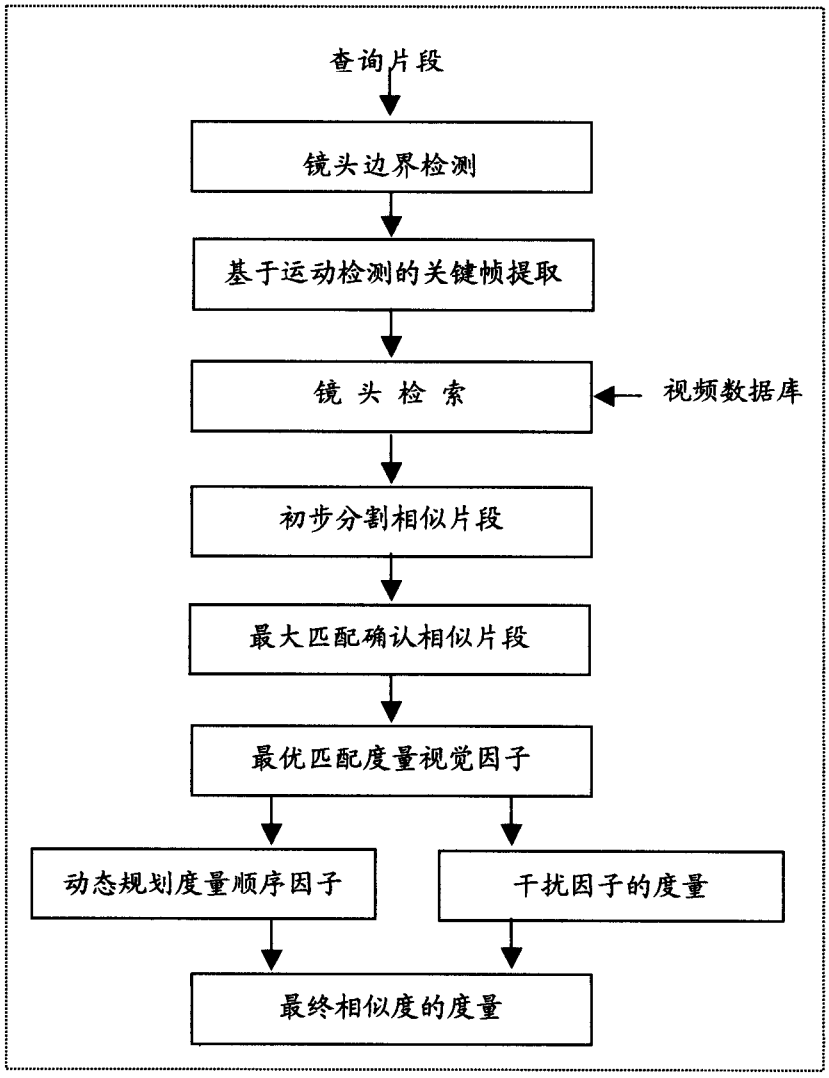

查询片段
镜头边界检测
基于运动检测的关键帧提取
镜 头 检 索
视频数据库
初步分割相似片段
最大匹配确认相似片段
最优匹配度量视觉因子
动态规划度量顺序因子
干扰因子的度量
最终相似度的度量

图 1

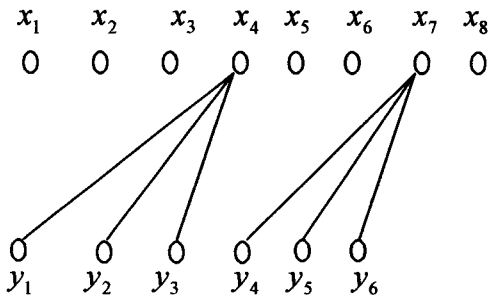

x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8
y_1 y_2 y_3 y_4 y_5 y_6

图 2

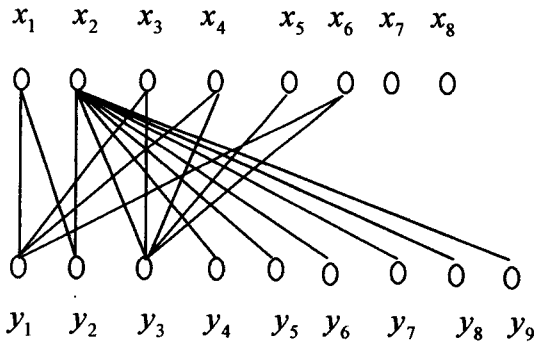

x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8
y_1 y_2 y_3 y_4 y_5 y_6 y_7 y_8 y_9

图 3

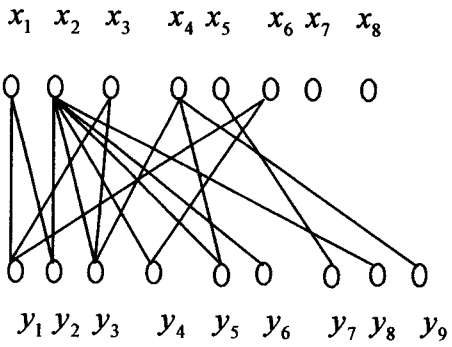

x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8
y_1 y_2 y_3 y_4 y_5 y_6 y_7 y_8 y_9

图 4

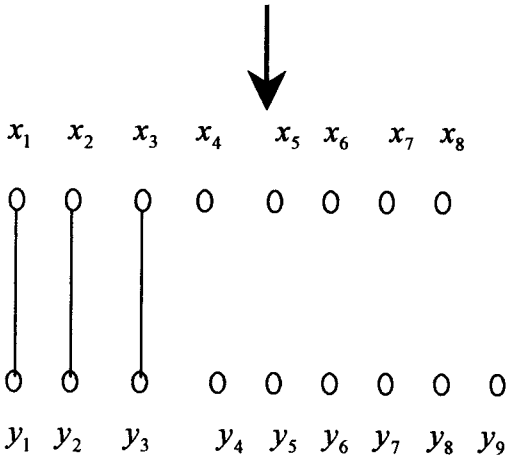

x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8
y_1 y_2 y_3 y_4 y_5 y_6 y_7 y_8 y_9

图 5

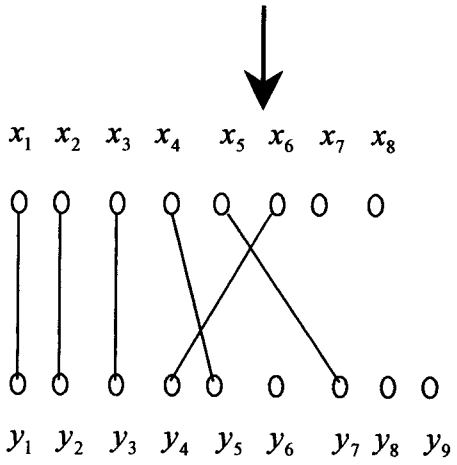

x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8
y_1 y_2 y_3 y_4 y_5 y_6 y_7 y_8 y_9

图 6


059064.bmp 059079.bmp 059099.bmp 059108.bmp 059124.bmp 059175.bmp 059231.bmp 059285.bmp
059064.bmp 059079.bmp 059099.bmp 059108.bmp 059124.bmp 059175.bmp 059231.bmp 059285.bmp
090773.bmp 090793.bmp 090816.bmp 090835.bmp 090876.bmp 090906.bmp 090918.bmp 090927.bmp
226667.bmp 226683.bmp 226697.bmp 226750.bmp 226806.bmp 226817.bmp 226847.bmp
137049.bmp 137098.bmp 137151.bmp 137163.bmp 137169.bmp 137221.bmp 137229.bmp 137247.bmp

图 7