

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G06F 17/30 (2006.01)



[12] 发明专利申请公布说明书

[21] 申请号 200710065180.1

[43] 公开日 2008年10月8日

[11] 公开号 CN 101281520A

[22] 申请日 2007.4.5

[21] 申请号 200710065180.1

[71] 申请人 中国科学院自动化研究所

地址 100080 北京市海淀区中关村东路95号

[72] 发明人 胡卫明 李华北

[74] 专利代理机构 中科专利商标代理有限责任公司

代理人 周国城

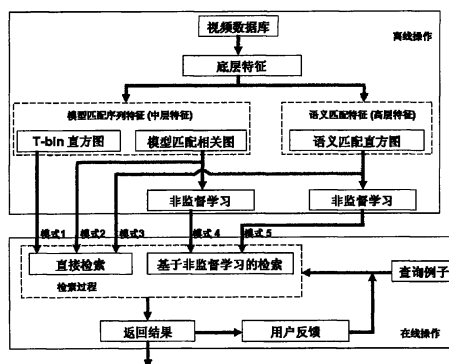
权利要求书4页 说明书14页 附图4页

[54] 发明名称

基于非监督学习和语义匹配特征交互式体育视频检索方法

[57] 摘要

本发明公开一种基于非监督学习和语义匹配特征的交互式视频检索方法，步骤包括：在视频数据库的视频图像帧层次提取图像底层特征、提取模型匹配序列特征；在图像底层特征的高级语义层次提取语义匹配特征；对提取模型匹配序列特征和语义匹配特征进行非监督学习，建立基于非监督学习的检索和直接检索，通过相关反馈形成交互界面。本发明集成视频的中层特征、高层特征、非监督检索机制和交互机制，构成一套新型完整的视频检索系统，精确的衡量了视频对象的时空序列信息，达到良好的检索效果，发展了对体育视频主题的语义理解，降低了系统的在线计算复杂度和检索时间，最后通过交互界面大幅度提高了系统的检索性能。



1、一种基于非监督学习和语义匹配特征的交互式体育视频检索方法，其特征在于，包括步骤如下：

步骤 1：在视频数据库的视频图像帧层次提取图像底层特征；

步骤 2：在图像底层特征的视频序列层次提取模型匹配序列特征；

步骤 3：在图像底层特征的高级语义层次提取语义匹配特征；

步骤 4：对提取模型匹配序列特征和语义匹配特征进行非监督学习，建立基于非监督学习的检索机制；

步骤 5：通过相关反馈技术形成交互检索界面，优化检索性能。

2、根据权利要求 1 所述的检索方法，其特征在于，模型匹配序列特征包括：加权 T-Bin 直方图和模型匹配相关图，用于反映视频对象的时空序列信息加权，T-Bin 直方图包括：每一维表示视频对象引用某一模型的频率，权重反映不同模型匹配序列的重要程度。

3、根据权利要求 1 所述的检索方法，其特征在于，模型匹配序列特征的提取步骤包括如下：

步骤 21：将整个数据库视为图像帧集合，对整个图像帧数据库进行下采样得到采样帧，将采样帧底层特征向量构造成矩阵形式生成训练集；

步骤 22：采用竞争学习算法学习得到模型集；

步骤 23：对视频对象中每一帧从模型集中找到 N 个最佳匹配模型，将视频的图像帧序列生成 N 条最佳匹配模型序列；

步骤 24：对 N 条最佳匹配模型序列提取加权 T-Bin 直方图和模型匹配相关图。

4、根据权利要求 2 所述的检索方法，其特征在于，模型匹配相关图包括：给定视频对象的最佳语义匹配序列为 S，任意两序列成员为 m_1 和 $m_2 \in S$ ，模型集为 MS，模型集所含模型数目为 Num_MS，像素跨度为 D；则该视频对象的模型匹配相关图为 Num_MS×D 维的向量；对于第 i 个模型 $Model_i \in MS$ 和某像素跨度 $k \in D$ ，模型匹配相关图的第 $(i-1) \times D + k$ 维如下所述：

$$MMC_{Model_i}^{(k)}(S) \square \prod_{m_1=Model_i} P [m_2 = Model_i \mid |m_1 - m_2| = k]$$

其物理意义为：对于某视频对象的模型匹配序列 S ，给定任一使用 $Model_i$ 的序列成员， $MMC_{Model_i}^{(k)}(S)$ 给出了 k 个像素跨度以外的序列成员也使用 $Model_i$ 的概率， MMC 同时描述视频对象的模型引用频率信息和序信息。

5、根据权利要求 1 所述的检索方法，其特征在于，语义匹配特征提取，包括如下步骤：

步骤 2a：选取具有代表性的带标注的视频对象组成小规模训练集，该训练集表征在当前底层特征描述能力的条件下数据库所涵盖运动主题；

步骤 2b：使用训练集，将 N 条最佳模型匹配序列进一步映射到语义标记层，得到了 N 条最佳标记序匹配列；

步骤 2c：对 N 条最佳标记匹配序列进行直方图提取并加权，得到视频对象的高层语义特征即语义匹配直方图。

6、根据权利要求 1 所述的检索方法，其特征在于，基于非监督学习的检索机制包括：采用主导集聚类算法对视频数据库进行非监督学习，将大部分相似度的计算转化为离线操作，用一致性函数来衡量每个生成的主导集的质量和制约总的主导集个数，其具体步骤包括：

步骤 31：将视频数据库作为无向边权图，其中每个视频对象作为无向边权图的结点，使用模型匹配相关图或语义匹配直方图，计算任意两段视频的相似度作为此结点对的权值，并生成全相似度矩阵 A ；

步骤 32：利用主导集聚类算法，取局部最优解中非零分量的标号集生成主导集；

步骤 33：将属于现有主导集的结点从当前图中删除，重复上述步骤直到结点图为空。

7、根据权利要求 1 所述的检索方法，其特征在于，交互式检索界面，用于对视频序列层次和语义层次实施相关反馈，包括如下：

采用最优查询相关反馈技术通过人机交互方式，用于帮助计算机了解用户需求，获得优化的查询向量，适用于直接检索机制；

采用关系矩阵相关反馈，用于调整各数据聚类间相互关系来涵盖数

据聚类间的全局语义关系，适用于基于非监督学习的检索机制；

语义校正相关反馈技术，用于对数据对象进行在线的语义校正、扩充数据库，适用于使用语义匹配直方图的检索过程。

8、根据权利要求7所述的检索方法，其特征在于：最优查询相关反馈技术，包括如下：用户对系统初始输出结果标记相关和不相关视频后，查询向量优化为：

$$f_q' = W_q \times f_q + W_R \times \left(\frac{1}{N_R} \sum f_R \right) - W_I \times \left(\frac{1}{N_I} \sum f_I \right)$$

式中： f_q 为原查询向量， f_R 、 f_I 、 N_R 、 N_I 为用户标记的相关视频和不相关视频及其个数， f_q' 为优化的查询向量， W_q 、 W_R 、 W_I 为常系数。

9、根据权利要求7所述的检索方法，其特征在于，关系矩阵相关反馈为如下三个步骤：

步骤 a：初始关系矩阵，计算任意两聚类中心间的相似度得到初始关系矩阵为：

$$Correlation_Matrix[i][j] = \exp(-1 * distance(Centroid_i, Centroid_j))$$

式中： $Centroid_i$ 、 $Centroid_j$ 为两聚类中心， $distance()$ 为某距离函数；

步骤 b：更新关系矩阵，关系向量 $F(x)$ 表示给定对象和各聚类中心的相似度为：

$$[F(x)]_i = \exp(-1 * distance(x, Centroid_i))$$

式中 x 为视频对象特征向量， $Centroid_i$ 为某聚类中心， $distance()$ 为某距离函数；

关系矩阵公式按下式进行更新：

$$Correlation_Matrix_k = Correlation_Matrix_{k-1} + \sum_{i=1}^{N_R} F(q)F(f_R) - \sum_{i=1}^{N_I} F(q)F(f_I)$$

式中： q 为查询向量， f_R 、 f_I 、 N_R 、 N_I 为用户标记的相关视频和不相关视频及其个数， k 为更新次数；

步骤 c：使用关系矩阵检索，对于某一查询请求，在关系矩阵中找到 N 个最相关聚类，再从中返回查询结果；在每次反馈后，相关矩阵的更新将被保存，使得性能提升得以累积。

10、根据权利要求7所述的检索方法，其特征在于，语义校正相关反馈技术的具体步骤如下：

步骤 d: 根据用户反馈意见, 得到相关视频集 RS 和不相关视频集 IS;

步骤 e: 对相关视频集 RS 和不相关视频集 IS 分别计算均值向量 RMV 和均值向量 IMV;

步骤 f: 在均值向量 RMV 中找出最大数值的两个分量 RD 和 RD2, 表示对应该视频最相关的两个主题;

步骤 g: 在均值向量 IMV 中找到最大数值的分量 ID, 表示最不相关的主题;

步骤 h: 如果 $(ID == RD1) \quad RD = RD2$, 执行步骤 i;

步骤 i: 优化查询的语义匹配直方图:

$Query_SMH[RD] = 1, \quad Query_SMH[ID] = 0;$

步骤 j: 将新的特征存入数据库, 重新检索。

基于非监督学习和语义匹配特征交互式体育视频检索方法

技术领域

本发明涉及计算机应用技术领域，特别涉及多媒体检索技术。

背景技术

随着多媒体技术和计算机网络的飞速发展，全世界包括数字图像、音频、视频在内的多媒体数据正在以惊人的速度增长。每天新生成的数千兆字节的多媒体数据，由于其分布的无序性，就如同泛滥的洪水般无情的淹没了其中所包含的大量有用信息。面对如此丰富且散布在世界范围内的海量多媒体资源，用户如何才能有效的利用信息及互联网新技术实现对所需多媒体资源的快捷定位，方便获取和有效管理已成为一个十分迫切问题，也使得多媒体检索技术逐渐成为一个非常活跃的研究领域。

基于内容的多媒体检索是指对多媒体数据所蕴含的物理内容和语义内容进行计算机分析和理解，以方便用户查询，其本质就是对无序的多媒体数据流结构化，提取语义信息，保证多媒体内容能够被快速检索。基于内容的视频检索和基于内容的图像检索是多媒体检索的两个最重要的分支。近年来，由于编码、计算机多媒体处理和网络传输技术飞速发展，用户可以通过高速互联网实时查询、欣赏和产生丰富多彩的视频数据，如电影、动画、新闻和体育节目等，并利用计算机自动处理视频数据流。视频已经成为人们传输和获取信息的主要途径之一。本质上，视频是由连续的图像帧序列组成的连续数据流，是二维数字图象和时间维度构成的三维对象。主要特点如下：(1) 视频数据较之于图像包含更多的视觉信息和语义信息；(2) 视频的数据量巨大；(3) 视频的抽象化、结构化程度低，对其进行有效的管理和方便的检索存在着很大的困难。因此，基于内容的视频检索已成为计算机应用技术和模式识别领域最热门的研究课题之一，具有十分广阔的应用前景。

视频数据本质上就是图像序列，视频检索中所用的视频特征也是由

图像帧的底层特征组成，同时图像检索中所用到的检索机制和思想也为视频检索的发展奠定了坚实的基础。因此图像检索的基本技术可以说是一个必修环节。基于内容的图像检索(CBIR, Content-based Image Retrieval)是指直接采用图像内容进行图像信息查询的检索技术。其主要思想是根据图像所包含的色彩、纹理、形状以及对象的空间关系等低层图像特征来分析图像信息,建立图像的特征向量作为索引;用户检索时一般是提供样例图像,系统抽取该样例图像的相应特征向量,然后同数据库中所有被检索对象的特征向量进行比较,并将与样例特征相似的图像返回给用户。

基于内容的视频检索主要是依赖于视频的视觉特征和时空特征。常用的检索方式是基于视频例子的检索,用户提交视频例子,检索系统从大型视频数据库中返回用户需要的相似视频。视频数据是由二维空间和时间构成的三维对象,如何定义视频之间的相似性量度是一个难点,需要解决以下关键问题:

- (1) 视频不是简单的帧序列集合,而是由场景—组—镜头—关键帧组成的层次结构,视频间相似度衡量在哪个层次上进行是视频比较的前提;
- (2) 关键帧的视觉特征是整部视频视觉特征的基础,但每部视频都有相当数量的关键帧;对于大型视频数据库而言,所有视频的各个关键帧视觉特征的存储量和相互间比较次数都是很可观的;
- (3) 两部视频是否相似是一个很复杂的问题,不同用户有不同的理解,掺杂着人为主观因素,要设计合理的视频比较算法必须尽量综合考虑各种因素。

基于镜头质心向量的视频检索系统引入镜头质心向量这一新概念,计算镜头层次的相似度进而又产生视频层次的相似度;利用关键帧间的数据冗余,通过牺牲一定的视频时空信息,大大降低了关键帧特征的存储量,简化了系统的复杂程度,实现了基于内容的视频检索的基本方法。

“iARM”系统使用基于模型的方法对视频的时空序列信息进行准确的建模并将对象内容映射到事先生成的模型上去。该系统强调对视频时空信息建模的准确性,因此其相关反馈分析仅仅需要有限次反馈和较少的训练样本即可获得较小的检索时间和较好的检索性能。

除上述技术外，基于图理论的非监督学习算法、图像检索相关反馈技术和基于用户反馈的信息嵌入技术等都为本发明奠定了基础。

发明内容

本发明的目的是提出新的中层、高层视频特征，以反映视频的序列时空信息和语义主题；建立新的基于非监督学习的检索机制，以降低相似度在线计算复杂度和减少检索时间；构建新的交互式检索界面，以在线优化查询向量、在线校正语义标记、提高检索性能和扩充数据库，为此，本发明提供一种基于非监督学习和语义匹配特征的交互式体育视频检索方法。

为实现上述目的，本发明提供基于非监督学习和语义匹配特征的交互式体育视频检索方法，包括步骤如下：

步骤 1：在视频数据库的视频图像帧层次提取图像底层特征；

步骤 2：在图像底层特征的视频序列层次提取模型匹配序列特征；

步骤 3：在图像底层特征的高级语义层次提取语义匹配特征；

步骤 4：对提取模型匹配序列特征和语义匹配特征进行非监督学习，建立基于非监督学习的检索机制；

步骤 5：通过相关反馈技术形成交互检索界面，优化检索性能。

根据本发明的实施例，模型匹配序列特征包括：加权 T-Bin 直方图和模型匹配相关图，用于反映视频对象的时空序列信息，加权 T-Bin 直方图包括：每一维表示视频对象引用某一模型的频率，权重反映不同模型匹配序列的重要程度。

根据本发明的实施例，模型匹配序列特征的提取步骤包括如下：

步骤 21：将整个数据库视为图像帧集合，对整个图像帧数据库进行下采样得到采样帧，将采样帧底层特征向量构造成矩阵形式生成训练集；

步骤 22：采用竞争学习算法学习得到模型集；

步骤 23：对视频对象中每一帧从模型集中找到 N 个最佳匹配模型，将视频的图像帧序列生成 N 条最佳匹配模型序列；

步骤 24：对 N 条最佳匹配模型序列提取加权 T-Bin 直方图和模型匹配相关图。

根据本发明的实施例，模型匹配相关图包括：给定视频对象的最佳语义匹配序列为 S ，任意两序列成员为 m_1 和 $m_2 \in S$ ，模型集为 MS ，模型集所含模型数目为 Num_MS ，像素跨度为 D ；则该视频对象的模型匹配相关图为 $Num_MS \times D$ 维的向量；对于第 i 个模型 $Model_i \in MS$ 和某像素跨度 $k \in D$ ，模型匹配相关图的第 $(i-1) \times D + k$ 维如下所述：

$$MMC_{Model_i}^{(k)}(S) \square P_{m_1=Model_i} [m_2 = Model_i \mid |m_1 - m_2| = k]$$

其物理意义为：对于某视频对象的模型匹配序列 S ，给定任一使用 $Model_i$ 的序列成员， $MMC_{Model_i}^{(k)}(S)$ 给出了 k 个像素跨度以外的序列成员也使用 $Model_i$ 的概率， MMC 同时描述视频对象的模型引用频率信息和序信息。

根据本发明的实施例，语义匹配特征提取，包括如下步骤：

步骤 2a：选取具有代表性的带标注的视频对象组成小规模训练集，该训练集表征在当前底层特征描述能力的条件下数据库所涵盖运动主题；

步骤 2b：使用训练集，将 N 条最佳模型匹配序列进一步映射到语义标记层，得到了 N 条最佳标记序匹配列；

步骤 2c：对 N 条最佳标记匹配序列进行直方图提取并加权，得到视频对象的高层语义特征即语义匹配直方图。

根据本发明的实施例，基于非监督学习的检索机制包括：采用主导集聚类算法对视频数据库进行非监督学习，将大部分相似度的计算转化为离线操作，用一致性函数来衡量每个生成的主导集的质量和制约总的主导集个数，其具体步骤包括：

步骤 31：将视频数据库作为无向边权图，其中每个视频对象作为无向边权图的结点，使用模型匹配相关图或语义匹配直方图，计算任意两段视频的相似度作为此结点对的权值，并生成全相似度矩阵 A ；

步骤 32：利用主导集聚类算法，取局部最优解中非零分量的标号集生成主导集；

步骤 33：将属于现有主导集的结点从当前图中删除，重复上述步骤直到结点图为空。

根据本发明的实施例，交互式检索界面，用于对视频序列层次和语义层次实施相关反馈，包括如下：采用最优查询相关反馈技术通过人机交互方式，用于帮助计算机了解用户需求，获得优化的查询向量，适用于直接检索机制；采用关系矩阵相关反馈，用于调整各数据聚类间相互关系来涵盖数据聚类间的全局语义关系，适用于基于非监督学习的检索机制；语义校正相关反馈技术，用于对数据对象进行在线的语义校正、扩充数据库，适用于使用语义匹配直方图的检索过程。

根据本发明的实施例，最优查询相关反馈技术，包括如下：用户对系统初始输出结果标记相关和不相关视频后，查询向量优化为：

$$f_q' = W_q \times f_q + W_R \times \left(\frac{1}{N_R} \sum f_R \right) - W_I \times \left(\frac{1}{N_I} \sum f_I \right)$$

式中： f_q 为原查询向量， f_R 、 f_I 、 N_R 、 N_I 为用户标记的相关视频和不相关视频及其个数， f_q' 为优化的查询向量， W_q 、 W_R 、 W_I 为常系数。

根据本发明的实施例，关系矩阵相关反馈为如下三个步骤：

步骤 a：初始关系矩阵，计算任意两聚类中心间的相似度得到初始关系矩阵为：

$$\text{Correlation_Matrix}[i][j] = \exp(-1 * \text{distance}(\text{Centroid_i}, \text{Centroid_j}))$$

式中： Centroid_i 、 Centroid_j 为两聚类中心， $\text{distance}()$ 为某距离函数；

步骤 b：更新关系矩阵，关系向量 $F(x)$ 表示给定对象和各聚类中心的相似度为：

$$[F(x)]_i = \exp(-1 * \text{distance}(x, \text{Centroid_i}))$$

式中 x 为视频对象特征向量， Centroid_i 为某聚类中心， $\text{distance}()$ 为某距离函数；

关系矩阵公式按下式进行更新：

$$\text{Correlation_Matrix}_k = \text{Correlation_Matrix}_{k-1} + \sum_{i=1}^{N_R} F(q)F(f_R) - \sum_{i=1}^{N_I} F(q)F(f_I)$$

式中： q 为查询向量， f_R 、 f_I 、 N_R 、 N_I 为用户标记的相关视频和不相关视频及其个数， k 为更新次数；

步骤 c：使用关系矩阵检索，对于某一查询请求，在关系矩阵中找

到 N 个最相关聚类，再从中返回查询结果；在每次反馈后，相关矩阵的更新将被保存，使得性能提升得以累积。

根据本发明的实施例，语义校正相关反馈技术的具体步骤如下：

步骤 d：根据用户反馈意见，得到相关视频集 RS 和不相关视频集 IS；

步骤 e：对相关视频集 RS 和不相关视频集 IS 分别计算均值向量 RMV 和均值向量 IMV；

步骤 f：在均值向量 RMV 中找出最大数值的两个分量 RD 和 RD2，表示对应该视频最相关的两个主题；

步骤 g：在均值向量 IMV 中找到最大数值的分量 ID，表示最不相关的主题；

步骤 h：如果 $(ID == RD1) \quad RD = RD2$ ，执行步骤 i；

步骤 i：优化查询的语义匹配直方图：

$$\text{Query_SMH}[RD] = 1, \text{Query_SMH}[ID] = 0;$$

步骤 j：将新的特征存入数据库，重新检索。

本发明集成了视频的中层特征、高层特征、非监督学习检索机制和交互机制，构成了一套新型完整的视频检索系统，精确的衡量了视频的时空序列信息，发展了对体育视频主题的语义理解，降低了系统的在线计算复杂度和检索时间，通过交互界面大幅度提高了系统的检索性能，具有广阔的应用前景。

附图说明

图 1 为本发明系统结构框图。

图 2 为模型匹配序列特征中视频的模式匹配示意图。

图 3 为语义匹配特征中视频的标记匹配示意图。

图 4 为关系矩阵相关反馈技术中数据聚类全局语义关系示意图。

图 5 为“CBVR_System”的程序界面示意图。

图 6 为使用 MMC 的非监督学习检索模式对于排球查询的返回结果。

图 7 为使用 SMH 的直接检索模式对于排球查询的返回结果。

图 8 为直接检索结果和经过一次反馈后的结果的比较。

具体实施方式

下面将结合附图对本发明加以详细说明，应指出的是，所描述的实施例仅旨在便于对本发明的理解，而对其不起任何限定作用。

本发明的整体框架见图 1。程序“CBVR_System”是实现本发明方法的一个具体实例，采用一台计算机，用 Visual C++编程实现。本发明提出的基于非监督学习和语义匹配特征的交互式体育视频检索方法主要包括如下四个关键问题：

- (1) 模型匹配序列特征；
- (2) 语义匹配特征；
- (3) 基于非监督学习的检索机制；
- (4) 交互式检索界面。

本发明总体结构可分为离线操作和在线操作两个部分。离线操作部分由特征提取和非监督学习组成。首先在视频图像帧层次对数据库对象提取图像底层特征；随后在视频序列层次提取模型匹配序列特征；在高级语义层次提取语义匹配特征；对所提取的视频特征进行非监督学习，建立基于非监督学习的检索机制；在线操作又分为检索机制和交互反馈。检索机制提供五种检索模式：使用 TBH 的直接检索、使用 MMC 的直接检索、使用 SMH 的直接检索、使用 MMC 的非监督学习检索和使用 SMH 的非监督学习检索；当收到查询请求时，系统按照用户选择模式进行检索并返回结果；最后通过相关反馈技术形成交互界面完善系统性能。下面详细给出该发明技术方案中所涉及各个细节问题的说明。

(1) 模型匹配序列特征

本发明中的模型匹配序列特征是由图像帧底层特征生成的视频中层特征，是在视频层次衡量时空信息的序列特征，又是依赖学习的基于模型的特征。其关键问题在于模型集的获取及视频中层序列特征的选择。模型集获取是序列特征提取的第一个关键，一个“好”模型表示了视频数据库内一帧极具代表性的广义图像，而“好”模型集就是由这样一组代表性强而彼此相关性弱的模型组成。获得适当模型集后，通过对各视频帧进行模型匹配，可将视频片断映射为一组最佳模型匹配序列。从最佳模

型匹配序列中提取怎样的中层序列特征成为第二个关键问题，不同的特征提取方法对相似度计算的影响巨大，从而很大程度上决定了检索的性能。

本发明中，对数据库提取底层特征之后，即可进行模型匹配序列特征的提取。模型匹配序列特征是通过模型匹配的方法由图像帧底层特征生成的视频中层序列特征，用以精确的反映视频对象的时空序列信息。描述图像视觉信息的底层特征是检索系统的基础，但其并非本发明关注的重点，采用何种底层特征不会影响本发明的结构及原理，所以本系统简单的采用了颜色相关图。具体提取过程主要由训练集生成、模型生成、模型匹配和特征提取四个步骤组成。

步骤一：训练集生成，将整个数据库视为图像帧集合而不是视频集合；通过对整个图像帧数据库进行下采样得到采样帧，将采样帧底层特征向量构造成矩阵形式作为训练集。

步骤二：模型生成，本发明采用竞争学习算法学习得到模型集。

首先，每次从训练集中随机选取一个样本 x_v ，按公式（1）为其在现有模型集 $\{m_t, t = 1 \cdots T\}$ 中找到最佳匹配模型 m_{t^*} ：

$$\|x_v - m_{t^*}\| < \|x_v - m_t\|, t = 1 \cdots T, t \neq t^* ; \quad (1),$$

随后，按公式（2）实施竞争学习算法，其中 m 为迭代次数， $l(m)$ 为随 m 递减的学习步长系数。当迭代次数 m 达到指定次数后，即得到模型集。

$$m_{t^*}(m+1) = m_{t^*}(m) + l(m)(x_v - m_{t^*}(m)); \quad (2),$$

步骤三：模型映射，具体模型映射过程见图 2：给定某视频帧序列对象，为其每一帧从模型集中找到 N 个最佳匹配模型。从而给定视频的图像帧序列就变为了 N 条最佳匹配模型序列。

步骤四：特征提取，在最佳匹配模型序列的基础上，本发明定义了加权 T-Bin 直方图（WTH）和模型匹配相关图（MMC）。

加权 T-Bin 直方图每一维表示某一模型被引用的频率，权重反映了不同模型匹配序列的重要程度。WTH 反映了相应模型被引用的频率信息但忽略了序列的先后顺序。

定义 1. 模型匹配相关图: S 为给定视频对象的最佳语义匹配序列, m_1 和 $m_2 \in S$ 为任意两序列成员, MS 为模型集, 模型集所含模型数目为 Num_MS , D 为像素跨度; 则该视频对象的模型匹配相关图定义为一个 $Num_MS \times D$ 维的向量; 对于第 i 个模型 $Model_i \in MS$ 和某像素跨度 $k \in D$, 模型匹配相关图的第 $(i-1) \times D + k$ 维如下定义:

$$MMC_{Model_i}^{(k)}(S) \square \prod_{m_1=Model_i} P [m_2 = Model_i \mid |m_1 - m_2| = k] \quad (3),$$

对于某视频对象的模型匹配序列 S , 给定任一使用 $Model_i$ 的序列成员, $MMC_{Model_i}^{(k)}(S)$ 表示给出了 k 距离以外的序列成员也使用 $Model_i$ 的概率。MMC 同时描述出视频对象的模型引用频率信息和引用序信息。

(2) 语义匹配特征

本发明定义了语义匹配特征即语义匹配直方图 (SMH)。体育视频可以简单的通过它们的体育项目名称对其主题进行索引, 如篮球、橄榄球和网球等。语义匹配直方图是在图像底层特征和视频中层特征的基础上, 通过模型匹配和主动学习的方法生成的视频高层语义特征, 用以对体育视频对象进行主题标注。SMH 给出某视频对象属于各个语义主题的概率, 并用最相关主题标记该视频。与中层特征相比, SMH 一定程度上反映出视频的语义内容, 降低了特征向量维度, 提高了检索性能。本发明开发出模型匹配主动学习算法用以提取语义特征。该算法通过对标记训练集进行主动学习得到一组带语义标记的模型集, 继而应用该模型集对视频对象进行模型匹配得到若干条最佳标记匹配序列, 再对匹配序列进行直方图提取, 最后得到语义匹配直方图 (SMH)。具体提取方法可分为如下三个步骤。

步骤一: 训练集, 训练集作为主动学习机制的监督信息来源, 本系统选取具有代表性的带标注的视频对象组成小规模训练集。也就是说, 训练集中的视频对象事先已人工标记了运动项目主题; 相对于数据库来说训练集规模较小; 在当前底层特征描述能力的条件下训练集可以涵盖尽可能多的运动主题。

步骤二: 模型生成和映射, 映射过程见图 3, 模型生成与模型匹配序列特征的对应步骤相似, 差别仅为使用了不同的训练集。(在图 2 中加

入语义标记层形成图 3), 将 N 条最佳模型匹配序列进一步映射到语义标记层, 得到了 N 条最佳标记序匹配列。

步骤三: 生成语义匹配直方图 (SMH), 对 N 条最佳标记匹配序列进行直方图提取并加权, 得到视频对象的高层语义特征——SMH。SMH 具有如下特点: 维度低, 相当于训练集中监督信息的语义主题数目; 物理意义明确, 其每一维表示视频对象属于某相应主题的概率, 并用最相关主题标记该对象; 向量稀疏, 大幅度减少存储空间和简化相似度计算。

(3) 基于非监督学习的检索

传统的检索方法是相似度排序机制, 该机制具有直接、灵活和易于实现相关反馈等特点; 但对于每一个查询对象, 该方法都要重新在线计算所有的相似度, 同时遍历整个数据空间, 因此需要承受很高的在线计算复杂度; 特别对于大型视频数据库而言, 该机制几乎无法正常工作。

本发明建立了基于非监督学习的检索框架用以代替传统的直接排序检索机制。该机制通过对视频数据库进行非监督学习, 使大部分相似度的计算转化为离线操作, 大幅度降低了在线计算的复杂度, 同时实现对数据库更加有效的管理。这样, 检索过程可分为粗检索和细检索两部分, 粗检索实际就是一个将查询分到已有聚类的分类过程, 而细检索则只需要在线计算少量的相似度 (查询和某聚类样本、查询和少量自由样本的相似度)。本部分关键问题在于非监督学习算法的选择。聚类算法是基于非监督学习检索机制的核心。聚类时间、聚类纯度和聚类个数严重影响着检索系统的性能。

本系统采用主导集聚类算法实现基于非监督学习的检索机制。主导集聚类算法 (Dominant Set Clustering) 是图理论聚类算法的一种, 该算法存在一致性函数来衡量每个生成的主导集的质量和制约总的主导集个数。与其他聚类算法相比, 主导集聚类产生的聚类具有较高纯度, 聚类个数通过一致性阈值的设定可以自动确定, 同时计算复杂度相对较小, 具体聚类算法可分为三个步骤。

步骤一: 全相似度矩阵, 将视频数据库视为无向边权图, 其中每个视频对象作为图的结点, 使用模型匹配相关图或语义匹配直方图, 计算任意两段视频的相似度作为此结点对的权值, 并组成全相似度矩阵 A。

步骤二：迭代方程，主导集聚类算法等价于如下二次优化问题：

$$\max f(u) = u^T Au \quad s.t. \quad u \in \Delta, \quad (4),$$

其中 $\Delta = \left\{ u \in R^N : u_i \geq 0 \text{ and } \sum_{i=1}^n u_i = 1 \right\}$ ，A 为全相似度矩阵 (5)，

该问题的局部最优解可以通过如下迭代方程得到

$$u_i(t+1) = u_i(t)(Au(t))_i / u(t)^T Au(t), \quad t \text{ 为迭代次数} \quad (6),$$

取局部最优解中非零分量的标号集生成主导集；

步骤三：将属于现有主导集的结点从当前图中删除，重复上述步骤直到结点图为空。

(4) 交互式检索界面

相关反馈就是通过人机交互，让用户在线帮助计算机理解对象的语义信息和人类的主观需要。视频的序列特性使用户反馈意见需要相对较长的时间，所以视频检索的相关反馈技术发展十分有限。减轻用户使用负担，如何通过最少的反馈获得最好的检索效果成为视频相关反馈的发展趋势。

在本系统中，使用模型匹配序列特征的检索模式由于其本身并没有涉及视频的任何语义内容，所以其缺乏对检索对象语义的理解，同时忽视了人类感知的主观性。类似地，使用语义匹配特征的检索模式可从监督训练集中获取一定语义信息，然而此信息有时极为有限以至于不能始终保证对体育视频主题的准确标注。为反映用户的个性需求、弥补语义鸿沟以及在线补充有监督信息，本发明建立了一套交互式检索界面，分别于视频序列层次和语义层次实施了三种相关反馈技术：最优查询相关反馈技术、关系矩阵相关反馈技术和语义校正相关反馈技术，其中：

最优查询相关反馈技术：通过 TBH 或 MMC 等特征表示的查询向量常常不能准确地描述出用户的真正需求，所以本发明应用最优查询相关反馈技术通过人机交互方式，帮助计算机了解用户需求获得优化的查询向量。

用户对于系统的初始检索结果进行简单的打分，标记出相关视频和

非相关视频。根据用户反馈意见，最优查询向量可由公式（7）得到。这样，用户可以帮助计算机更加准确理解查询要求，提高了检索的性能。

$$f_q' = W_q \times f_q + W_R \times \left(\frac{1}{N_R} \sum f_R \right) - W_I \times \left(\frac{1}{N_I} \sum f_I \right) \quad (7)$$

式中： f_q 为原查询向量， f_R 、 f_I 、 N_R 、 N_I 为用户标记的相关视频和不相关视频及其个数， f_q' 为优化的查询向量， W_q 、 W_R 、 W_I 为常系数。

关系矩阵相关反馈技术：最优查询机制仅仅优化指定查询而忽略了整个数据库，所以通过交互而得到的性能提升总是在下一次检索时丢失，而得不到持续积累。即对同一查询要求再次检索时，还要重复全部交互过程。

有鉴于此，本发明提出了关系矩阵相关反馈技术，通过调整各数据聚类间相互关系来涵盖数据聚类间的全局语义关系见图4，图4中， C_1 - C_N 表示数据库中的 N 个聚类中心，权重 W 表示各聚类间的相似关系。具体过程分为如下三个步骤：

步骤一：初始关系矩阵。计算任意两聚类中心间的相似度得到初始关系矩阵：

$$Correlation_Matrix[i][j] = \exp(-1 * distance(Centroid_i, Centroid_j)) \quad (8)$$

式中： $Centroid_i$ 、 $Centroid_j$ 为两聚类中心， $distance()$ 为某距离函数。 $Correlation_Matrix$ 表示各聚类之间的相似关系。

步骤二：更新关系矩阵。关系向量 $F(x)$ 表示给定对象和各聚类中心的相似度：

$$[F(x)]_i = \exp(-1 \times distance(x, Centroid_i)) \quad (9)$$

式中 x 为视频对象特征向量， $Centroid_i$ 为某聚类中心， $distance()$ 为某距离函数。

关系矩阵按公式（10）进行更新：

$$Correlation_Matrix_k = Correlation_Matrix_{k-1} + \sum_{i=1}^{N_R} F(q)F(f_R) - \sum_{i=1}^{N_I} F(q)F(f_I) \quad (10)$$

式中： q 为查询向量， f_R 、 f_I 、 N_R 、 N_I 为用户标记的相关视频和不相关视频及其个数， k 为更新次数。

矩阵 $\sum_{i=1}^{N_r} F(q)F(f_{R_i})^T$ 中的非零分量表示用户认为较相关的聚类对；
同理 $\sum_{i=1}^{N_i} F(q)F(f_{I_i})^T$ 中的非零分量表示用户认为不相关的聚类对；通过加强相关聚类对的相似关系、减弱不相关聚类对的相似关系，关系矩阵得以更新。

步骤三：使用关系矩阵检索。对于某一查询请求，在关系矩阵中找到 N 个最相关聚类，再从中返回查询结果。在每次反馈后，相关矩阵的更新将被保存，使得性能提升得以累积。

语义校正相关反馈技术：用于对数据对象进行在线的语义校正、提高检索性能同时扩充数据库，具体算法如下：

步骤一：根据用户反馈意见，得到相关视频集 RS 和不相关视频集 IS；

步骤二：对相关视频集 RS 和不相关视频集 IS 分别计算均值向量 RMV 和均值向量 IMV；

步骤三：在 RMV 中找出最大数值的两个分量 RD 和分量 RD2，表示对应该视频最相关的两个主题；

步骤四：在 IMV 中找到最大数值的分量 ID，表示最不相关的主题；

步骤五：If (ID == RD1) RD = RD2；

步骤六：优化查询的语义匹配直方图：

Query_SMH[RD] = 1, Query_SMH[ID] = 0；

步骤七：将新的特征存入数据库，重新检索。

图 5—图 8 示出本发明的技术效果，其中：

图 5 为“CBVR_System”的程序界面示意图。其中，右上显示区显示数据库中视频的第一帧图像，使用按钮翻页，图片下方滑动条用于接收反馈信息；左上播放区用于播放选中视频；单选框用于选择检索模式；按钮区用于功能操作；状态栏实时显示程序运行信息。

图 6 为使用 MMC 的非监督学习检索模式对于排球查询的返回结果。使用单选框选择“使用 MMC 的非监督学习检索模式”，显示区输出检索结果的第一帧图像，单击图像可在播放区播放整段视频。状态栏显示检

索耗时。

图 7 为使用 SMH 的直接检索模式对于排球查询的返回结果。使用单选框选择“使用 SMH 的直接检索模式”，状态栏显示检索耗时和查询对象的语义主题。

图 8 为直接检索结果和经过一次反馈后的结果的比较。上图为“使用 THB 的直接检索模式”的原始输出结果，用户使用滑动条提供反馈意见（向右表示相关视频，向左表示不相关视频）；下图为反馈后结果，性能明显提高。

以上所述，仅为本发明中的具体实施方式，但本发明的保护范围并不局限于此，任何熟悉该技术的人在本发明所揭露的技术范围内，可理解想到的变换或替换，都应涵盖在本发明的包含范围之内，因此，本发明的保护范围应该以权利要求书的保护范围为准。

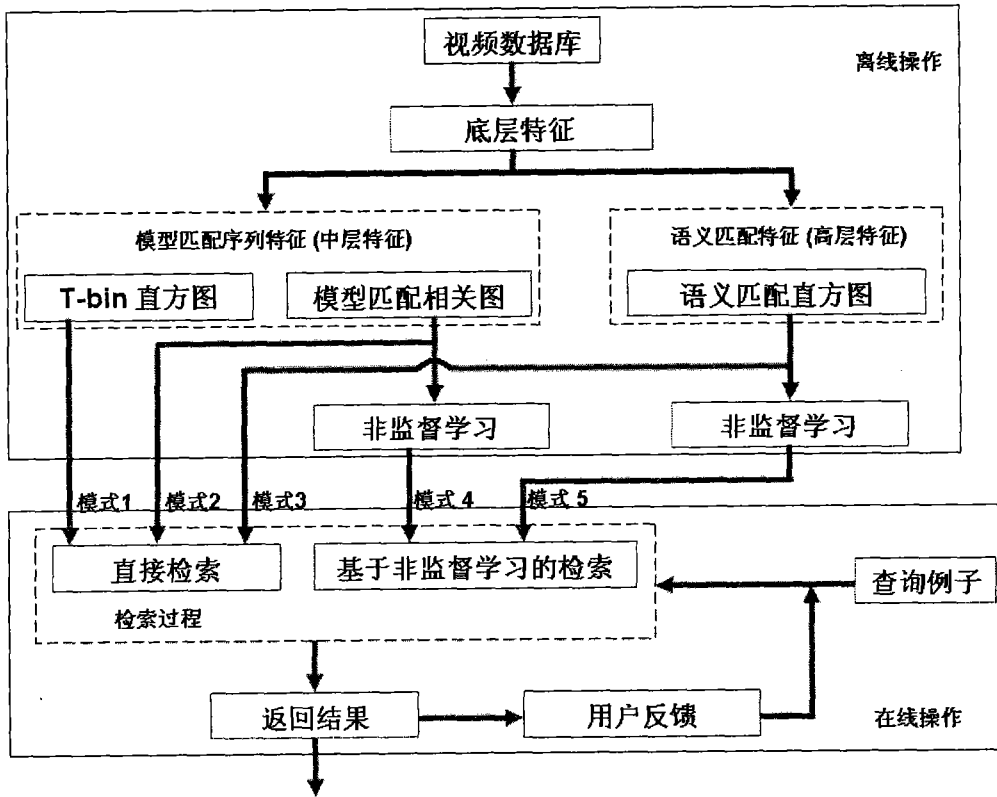


图 1

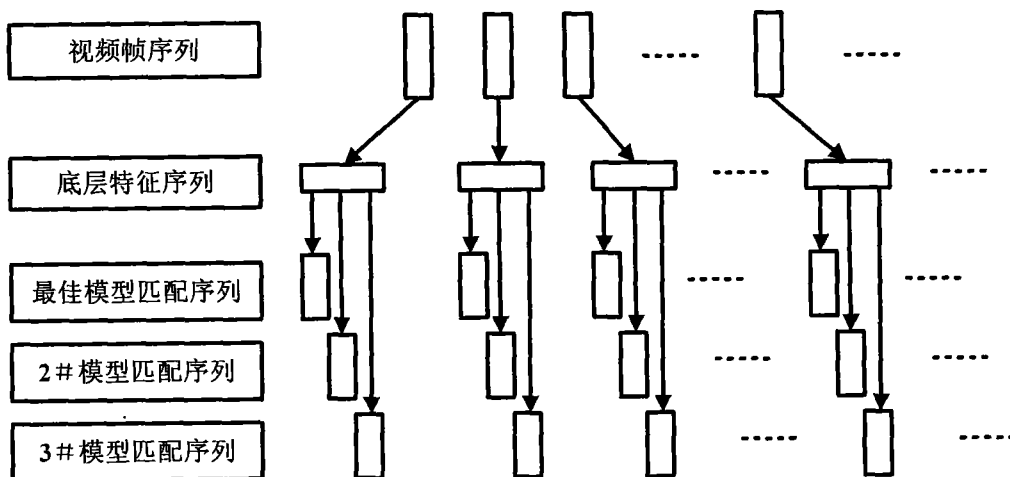


图 2

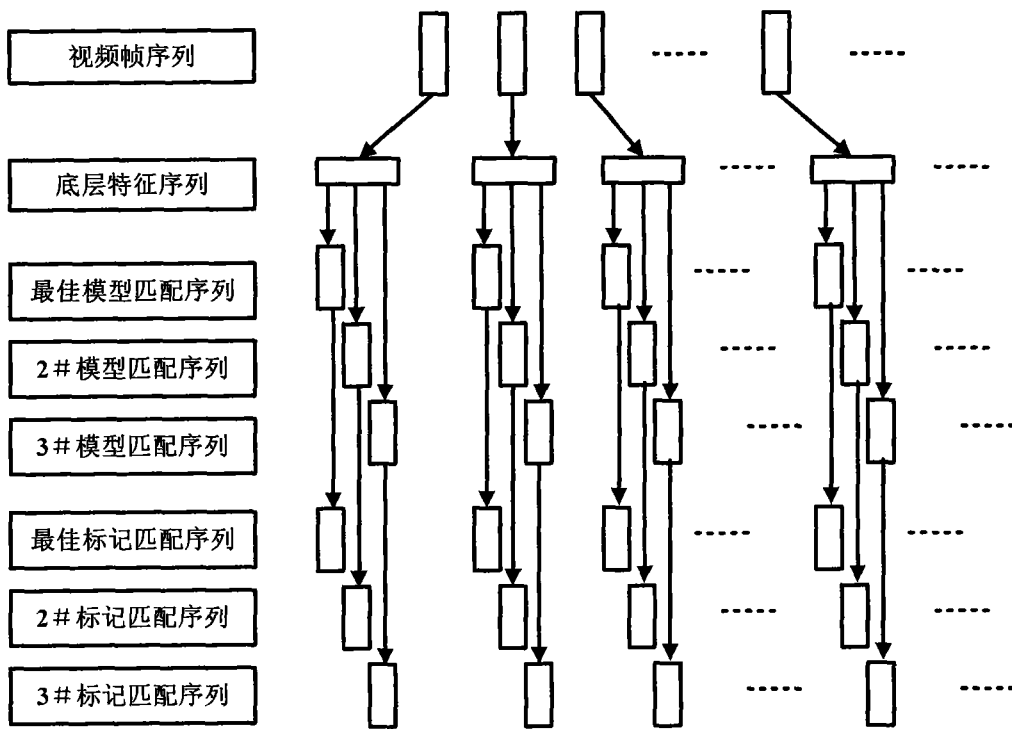


图 3

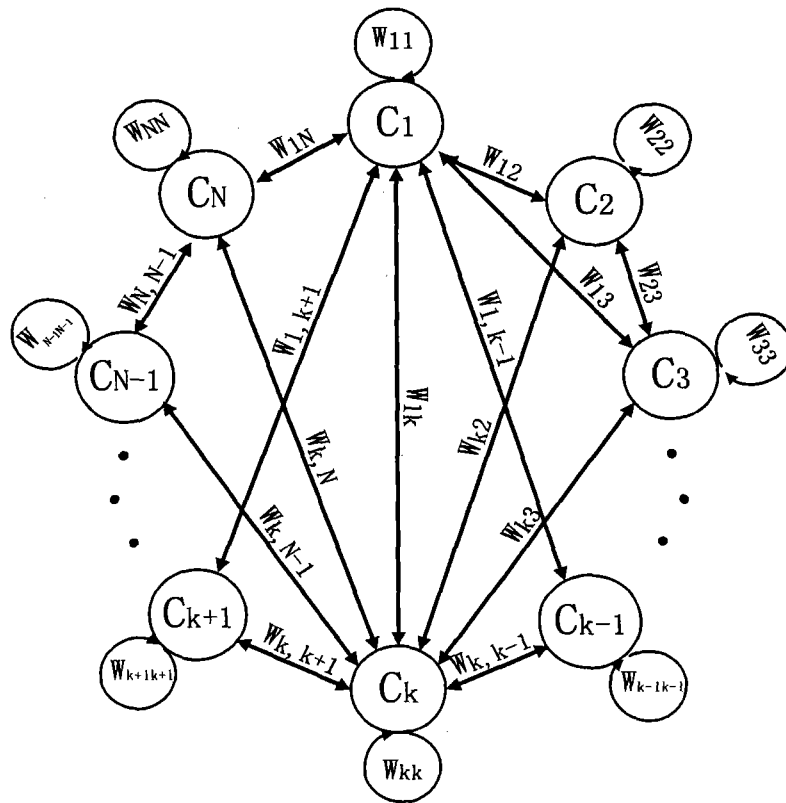


图 4

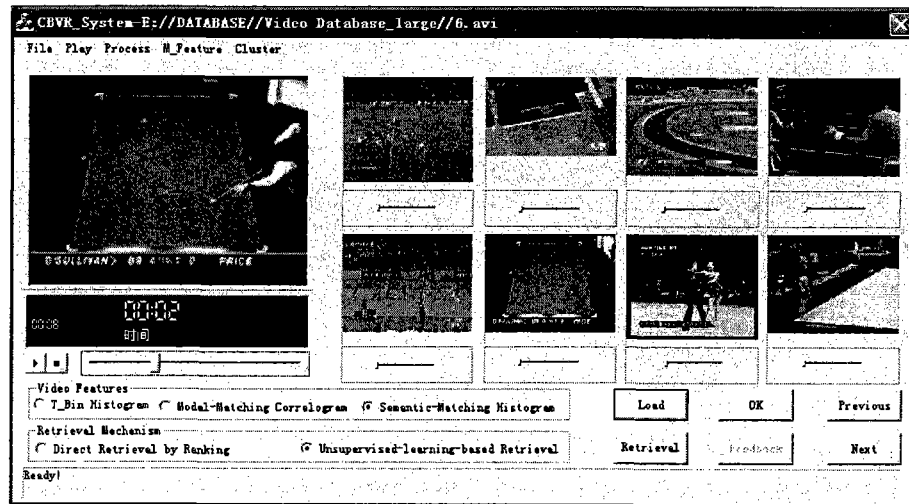


图 5

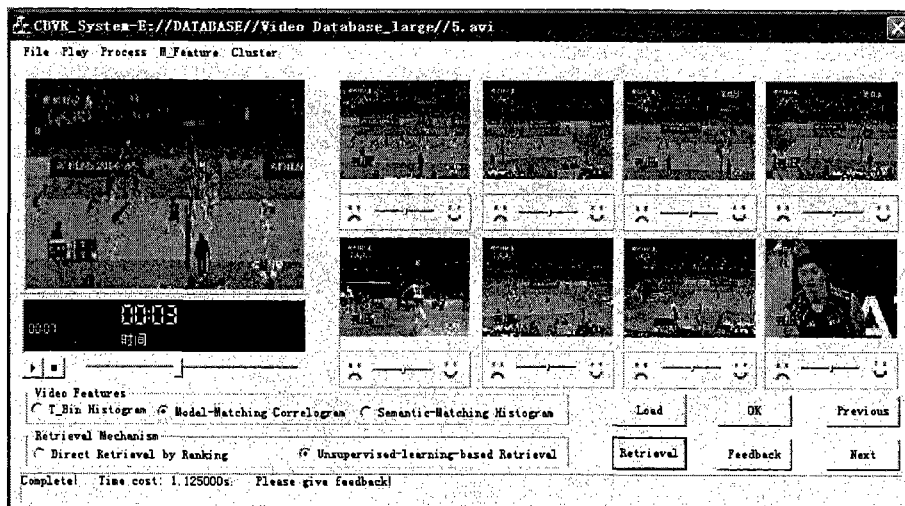


图 6

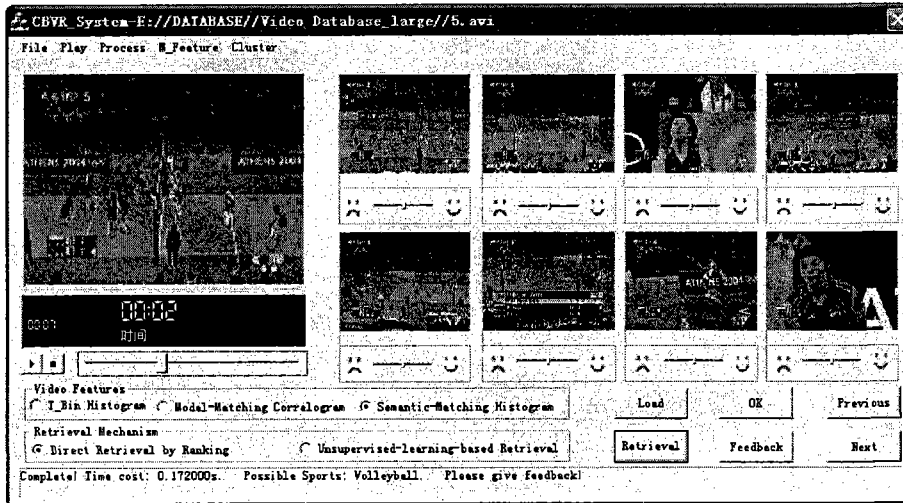


图 7

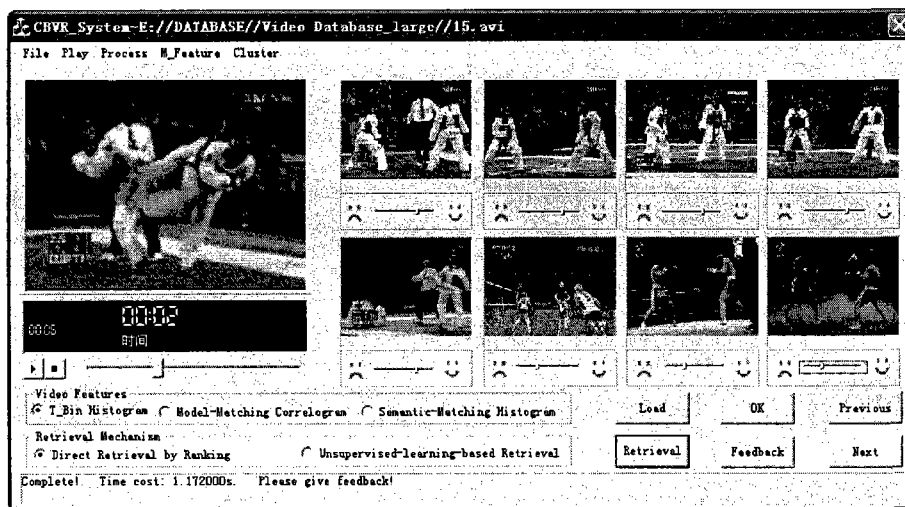


图 8