

交互式关键视频对象选择模型*

欧阳建权^{1,2,3+}, 李锦涛¹, 张勇东¹

¹(中国科学院 计算技术研究所,北京 100080)

²(湘潭大学 信息工程学院,湖南 湘潭 411105)

³(中国科学院 研究生院,北京 100049)

An Interactive Key Video Object Selection Model

OUYANG Jian-Quan^{1,2,3+}, LI Jin-Tao¹, ZHANG Yong-Dong¹

¹(Institute of Computing Technology, The Chinese Academy of Sciences, Beijing 100080, China)

²(College of Information Engineering, Xiangtan University, Xiangtan 411105, China)

³(Graduate School, The Chinese Academy of Sciences, Beijing 100049, China)

+ Corresponding author: Fax: +86-732-8292201, E-mail: oyjq@ict.ac.cn, <http://www.ict.ac.cn>

Ouyang JQ, Li JT, Zhang YD. An interactive key video object selection model. *Journal of Software*, 2007,18(3):547-554. <http://www.jos.org.cn/1000-9825/18/547.htm>

Abstract: Summarization of video can well represent the content of video. In the MPEG-1/2 standard key frame is used for representing the content of a video sequence. Similarly, in an object-based framework of the suggested MPEG-4 standard, key video object can summarize the content of video objects. In this paper, an interactive key video objects selection model (IKVOS) is presented as the result of improving the model of key video objects selection (KVOS). The model of KVOS is proved to satisfy the criterion of induction. Also, the course of IKVOS is formalized and the model of IKVOS is proved to satisfy the criterion of coinduction. It can dynamically generate key video objects with user's preference. The experimental results are given, as well as the evaluation of the performance of the proposed method whose distortion rate is lower than the current methods.

Key words: key video object; interactive key video object selection; interactive computing; induction; coinduction

摘要: 视频摘要能够充分表达视频信息.在 MPEG-1/2 中,关键帧可以反映镜头的内容.类似地,关键视频对象可以表示基于 MPEG-4 对象框架的视频内容.将关键视频对象选择模型(key video object selection model,简称 KVOS)扩展为基于交互计算的交互式关键视频对象选择模型(interactive key video object selection model,简称 IKVOS),并证明了关键视频对象选择模型满足归纳法准则,交互式关键视频对象选择模型满足余归纳法准则.同时用户可以根据偏好灵活选取关键视频对象.实验表明,根据紧凑和全面策略选择的关键视频对象,其失真率均低于相关方法.

关键词: 关键视频对象;交互式关键视频对象选择;交互计算;归纳法;余归纳法

* Supported by the National Natural Science Foundation of China under Grant No.60473002 (国家自然科学基金); the Key Project of Beijing Natural Science Foundation of China under Grant No.4051004 (北京市自然科学基金重点项目); the Scientific Research Fund of Hu'nan Provincial Education Department of China under Grant No.06B095 (湖南省教育厅科研项目)

Received 2004-12-30; Accepted 2006-04-17

中图法分类号: TP391

文献标识码: A

多媒体信息分析与检索是当前信息领域研究的热点之一。随着数字化技术的发展,数字多媒体资源,特别是视频产品越来越丰富,多媒体具有蕴涵信息量大、直观和印象深刻等优点,已经逐渐成为网络和个人电脑上重要的信息载体。传统的基于文本的信息检索技术很难对多媒体信息进行有效的结构化分析和快速检索,而多媒体信息分析与检索通过对多媒体从低层特征到高层语义进行处理、分析,获取其内容并根据内容进行检索。由此可见,多媒体信息分析与检索具有重要的理论意义和应用价值。视频摘要是对视频内容的概括,能够简洁地表现视频的重要内容,目的在于帮助人们更快捷、更方便地定位视频的重要内容,是多媒体信息分析与检索研究中的重要研究工作。目前,在基于帧的数字视频索引与检索模型中,主要是基于关键帧(代表帧)表示视频序列,虽然关键帧可以反映镜头的内容,但是不能提供该帧中的对象在镜头中的准确描述,而 MPEG-4 编码标准提供了基于对象的视频表示,视频对象(video object)在位流中单独编码,因此,与关键帧类似,基于关键视频对象(key video object,简称 KVO)的视频检索同样可以表示基于对象框架的视频摘要等主要内容。因此,在 MPEG-4 基于对象的浏览和检索框架中,关键视频对象将取代关键帧。由于关键视频对象可以紧致表示一个视频序列,因此可以应用于基于对象的视频浏览、检索与分析及视频监控等领域。

已有一些关于关键视频对象选择的研究。张宏江^[1]1997 年首次提出了基于关键对象的检索和浏览框架,但是,文中视频对象定义为连贯的运动区域,因此并未详细阐述对象提取过程。Ferman^[2]提出基于编码模型的关键视频对象平面(video object plane,简称 VOP)选择算法,主要运用内部编码宏块变化的百分比作为视频对象显著变化的标志,但是算法精度不够。Erol^[3]指出在其算法实验中,在所选择的 VOP 中,冗余 2 个 VOP,漏选了 1 个 VOP。Erol^[3]提出基于形状的自动关键视频对象平面选择算法,运用形状估计,实现基于修正的 Hamming 距离和 Hausdorff 距离的方法,在 MPEG-4 压缩域进行 VOP 选择,效率较高,不需要解码,比非压缩域的方法差不多快了 16×16 倍。Kim^[4]实现了基于关键对象的视频监控框架,其关键对象提取选用了基于区域和基于轮廓的形状描述子,基于区域的形状提取集成于视频监控系统。实验结果显示,该在线关键视频提取机制取得了较好的效果。Song^[5]通过挖掘关键对象选择和关键帧的内在关系,提出了高斯混合模型(Gaussian mixture model),较好地结合视频对象提取进行关键帧选择。

但是,由于关键视频对象选取有一定的主观性,以上方法不能根据用户偏好动态产生基于对象的视频摘要,因此是一个封闭系统。同时,上述方法的权值和阈值调整等都是基于经验的,也没有说明其经验值的合理性以及如何有效地确定其值等。因此,本文引入交互计算模型,提出一种基于交互计算的关键视频对象选择模型,即根据用户选择紧凑和全面策略可以灵活地产生关键视频对象集合。实验表明,根据紧凑和全面策略不仅可以灵活地选取关键视频对象,而且其失真率均低于相关方法,因此,交互式关键视频对象选择模型是对关键视频对象选择模型的扩展。

本文第 1 节介绍交互计算模型和交互式关键视频对象选择模型,两个模型分别满足归纳法和余归纳法。第 2 节为相应的实验结果与分析。第 3 节对全文工作进行总结,并提出工作展望。

1 基于交互计算的关键视频对象选择

1.1 交互计算模型概述

Wegner 和 Goldin 提出了一个计算模型——交互计算^[6],对经典的图灵机模型进行了扩展。在典型的图灵机模型中,当输入被确定以后,所有的计算过程都是确定的,并且相应的计算结果也是确定的。而交互计算的机器模型,即时序交互机(SIM)^[6]包含具有不确定性的 I/O 行为,即当前的输出对下一个输入产生影响,表达能力比典型的图灵机模型要强。交互计算模型已经在面向对象程序设计、分布式系统和软件工程等领域得到了应用。

而交互计算的数学模型是从算法到交互的计算表示的扩展,具体体现在一系列的代数扩展^[7]上:从归纳(inductive)到余归纳(coinductive)的扩展、从良基集(well-founded set)到非良基集(non-well-founded set)、从代数

(algebras)到余代数(coalgebras)的扩展.从归纳到余归纳的扩展表达了从字符串到流的转变,即从算法转变到交互的基础;非良基集主要用于作为形式化流的顺序行为,而良基集描述基于字符串的操作,比如程序终止证明等;余代数则为流的演算提供了工具,在交互计算模型中的作用相当于 λ 演算在 Church-Turing 模型中的作用^[7].下面介绍交互计算模型中的一些基本定义.

定义 1(代数)^[7]. 代数是一个结构 $A=(S,m:F(S)\rightarrow S)$,其中, S 是集合, $m:F(S)\rightarrow S$ 是一个保值的同态映射,按照句法规则构成的表达式集合映射到值的集合.

定义 2(余代数,coalgebra)^[8]. 余代数是一个结构 $CA=(S,m:S\rightarrow G(S))$,其中, S 是一个状态集, $m:S\rightarrow G(S)$ 是一个保行为(behavior)的同态映射,将状态未知的被观测系统映射到一组行为(观察).

由代数和余代数的定义可以看出,代数中结构的表达式是归纳定义的,演算过程也是归纳的.从代数到余代数的扩展就类似于从归纳到余归纳、从良基集到非良基集的扩展.

定义 3(交互性质)^[6]. 如果一个计算主体具有与不受它控制的外部环境交互的输入和输出动作,则它具有交互性质.

定义 4(归纳法)^[8]. 归纳法是从初始元素 x 通过迭代规则 $f(x)\rightarrow y$ 构造新的元素 y 的过程.该迭代条件 f 决定了一个传递闭包,满足 $f(x)=x$ 的最小不动点是该传递闭包的不动点.

归纳法满足:(1) 初始条件;(2) 基于最小不动点的迭代规则;(3) 最小化条件.

定义 5(F 余代数(F 系统))^[7]. F 代数是偶对 (A,α_A) ,其中, A 是集合, α_A 是函数: $F(A)\rightarrow A$.

定义 6(余归纳法,coinductive)^[8]. 令 S 为集合, (S,α_S) 是最终的 F 余代数.给定一个变换 $\alpha:S\rightarrow F(S)$,则存在唯一的同态映射 $\alpha_S:S\rightarrow F(S)$,我们说函数 f_α 由 α 通过余归纳定义.

余归纳法满足:(1) 基于最大不动点的迭代规则(即循环(circularity)条件);(2) 最大化条件.

由归纳法和余归纳法的定义可以看出:

- (1) 与归纳法相比,余归纳法去掉了初始条件,修改了迭代条件使之成为循环条件,用最大化条件替换最小化条件.
- (2) 余归纳法去掉静态的初始化,通过交互进行动态绑定规则,体现了从封闭系统到开放系统的扩展.
- (3) 归纳法的最小化条件只有规则允许的元素可以被包含,而余归纳法的最大化条件则刚好相反,所有规则没有禁止的元素都可以被包含.
- (4) 归纳法从本质上来讲是封闭的,而余归纳法从本质上来讲是开放的.

1.2 关键视频对象选择模型

一般来说,关键视频对象的选择应该反映视频对象的特征,如颜色、纹理和形状等内容的显著变化,而刻画这种变化则可以通过基于视频对象特征的相似度量函数来表示.因此,根据代数定义(定义 1),基于特征的关键视频对象选择方法可以形式化为以下模型:

定义 7. 关键视频对象选择模型(key video object selection model,简称 KVOS)可以用二元组 $(O,m:F(O)\rightarrow O)$ 描述,其中, O 是视频对象集合, F 是视频对象特征向量, m 是视频对象特征向量到关键视频对象集合的映射.

显然,基于相似度量函数的匹配规则可以看成是句法规则构成的表达式集合,因此, $m:F(O)\rightarrow O$ 是一个保值的同态映射,即关键视频对象选择模型满足代数的定义.

1.3 交互式关键视频对象选择模型

从信息论的角度来看,相关性小的视频对象帧比内容类似的视频对象帧表达更多的信息.因此,关键视频对象选择的首要标准是考虑视频对象之间的内容不相关性.

同时,关键视频对象反映了视频的重要内容.按照信息论的观点,通过选择关键视频对象,损失的信息应该越少越好.因此,关键视频对象选择的另一个重要标准是关键视频对象能够准确地表达视频信息.

本文将这两个标准形式化为最短路径问题和最小失真率问题,其定义如下:

定义 8. 有向图 $\{V,E\}$,其中 V 是 n 个视频对象组成的顶点集合 $\{v_0,\dots,v_{n-1}\}$, E 为带权值的边集合, $E\subseteq V\times V$,权

值 $w(v_i, v_j)$ 定义为两帧视频对象间的距离函数. 顶点 v_i 到顶点 v_m 路径为 $\{(v_i, v_j), (v_j, v_k), \dots, (v_l, v_m)\}$, 带权值的路径为 $\Omega(p) = \sum_{i=1}^k w(v_{i-1}, v_i)$. 关键视频对象选择可以转化为从顶点 v_0 到顶点 v_{n-1} 的最短路径问题 $\min \Omega(p)$.

定义 9. 失真率(distortion rate) $D(O) = \sum_{i=0}^{n-1} d(O_i, O'_i)$. 其中, $\{O_i\}$ 是视频对象集合, $i=0, \dots, n-1$, $\{kO_j\}$ 是关键视频对象集合 $j=0, \dots, m-1$, $KO_0=O_0$, $\{O'_i\}$ 是将关键对象替换原视频对象得到的集合, $O'_0 = KO_0$, $O'_i = O_{i=KO_j, i \leq KO_j}$ ($j=0, \dots, m-1, i=0, \dots, n-1$), $d(O_i, O'_i)$ 为 O_i 与 O'_i 之间的距离函数. 故关键视频对象选择可以转化为最小化失真率 $\min D(O)$.

本文首先提出基于最短路径约束的最小化失真率方法选择关键视频对象, 其原理为将 n 个视频对象组成的集合 $O = \{(k, id_k) | k \in \{0, \dots, n-1\}, id_k \in \{+1, -1\}\}$ 分为两类: 一类为关键视频对象集合 KO (标志 id_k 为 +1), 一类为非关键视频对象集合 $NonKO$ (标志 id_k 为 -1). 第 k 个视频对象对应的特征为 F_k . 步骤如下:

1. 初始化: for $i=0$ to $n-1$ $id_i = -1$; $kO_j = j = i = 0$ (所有视频对象认为是非关键视频对象).
2. $i=i+1$, if $i-kO_j > T_j$ (设定为帧率的 1/6) and $D(F_i, F_{ko}) \geq T_d$ (实验中设为 0.2) and $D(F_i, F_{ko_j}) \leq \lambda_{ko_j}$ ($D(F_i, F_{ko_j})$ 为两帧间视频对象的距离, $\lambda_{ko_j} = D(F_{ko_{j-3}}, F_{ko_{j-2}}) + D(F_{ko_{j-1}}, F_{ko_j})$ 为基于最短路径的估计值), then $id_i = +1, j = j+1, kO_j = i$.

repeat 2 until $i=n-1$.

A*算法^[9]实际上是一种启发式搜索, 所谓启发式搜索, 就是利用一个估价函数评估每次决策的价值, 决定先尝试哪一种方案. 如果一个估价函数能够在有限步内终止并找到最优解, 则称为可采纳性. A*算法是一个可采纳的最好优先算法. 算法的估价函数可表示为 $f(n) = g(n) + h(n)$, 其中, $f(n)$ 是节点 n 的估价函数, $g(n)$ 是起点到终点的最短路径值, $h(n)$ 是 n 到目标的最短路径的启发值.

为此, 本文提出基于最小化失真率约束的最短路径方法:

1. 初始化 OPEN 和 CLOSED 列表, $lastkey = i = 0$.
2. $i=i+1$, if $(g(v_i) \geq T_{\min})$ (T_{\min} 在实验中设为 0.18), then

$$g(v_i) = D(F_i, F_{lastkey}), h(v_i) = \frac{\sum_{j=lastkey+1}^i D(F_j, F_{j-1})}{i - lastkey}, f(v_i) = g(v_i) + h(v_i).$$

3. 基于 A*算法求出最短路径, 生成关键视频对象集合.

由于最短路径问题反映了关键视频对象选择的信息冗余度最小, 而最小失真率问题反映了关键视频对象选择的损失的信息最少. 这两者之间既存在冲突, 也存在一定的联系. 为此, 我们若考虑关键视频对象选择的“全面”, 则以最小失真率问题为基础、最短路径问题为约束, 就可以构建体现“全面”策略的关键视频对象选择方法, 即本文提出的基于最短路径约束的最小化失真率方法. 同时, 基于最短路径约束的最小化失真率方法可以看成是一种行为 (behavior), 该行为反映了关键视频对象选择的“全面”策略; 相应地, 若考虑关键视频对象选择的“紧凑”, 则以最短路径问题为基础, 最小失真率问题为约束, 就可以构建体现“紧凑”策略的关键视频对象选择方法, 即本文提出的基于最小化失真率约束的最短路径方法. 其中, 基于最短路径约束的最小化失真率方法和基于最短路径约束的最小化失真率方法体现了紧凑策略和全面策略的相互约束关系. 因此, 用户根据自己的偏好, 通过交互选择这些策略得到满意的关键视频对象集合.

由于关键视频对象选择有一定的主观性, 关键视频对象选择模型不能根据用户偏好动态产生基于对象的视频摘要, 因此是一个封闭系统. 为此, 本文将该模型扩展为基于交互计算的关键视频对象选择模型.

定义 10. 交互式关键视频对象选择模型 (interactive key video object selection model, 简称 IKVOS) 为二元组 $(S, \alpha_S: O \rightarrow O^S)$, 其中, S 是一个状态集 (由用户选择这两种策略), $\alpha_S: O \rightarrow O^S$ 是一个保行为的同态映射, 将状态未知的视频对象集合映射到一组行为 (包括基于最短路径约束的最小化失真率方法和基于最小化失真率约束的最短路径方法). 其中, O 为视频对象集合, O^S 是关键视频对象集合.

显然, 交互式关键视频对象选择模型满足余代数定义.

为此,我们首先证明关键视频对象选择模型满足归纳法准则.

命题 1. 关键视频对象选择模型满足归纳法准则.

证明:

1. 由于初始元素 $x(x$ 为视频对象)可以是第 1 个视频对象,关键视频对象选择规则为通过迭代规则 $f(x)(f(x)$ 是基于视频对象特征的相似度量函数,或者根据先验知识建模构造的函数) $\rightarrow y$ 构造新的关键视频对象 y ,故基于句法规则(可以将基于相似度量函数的匹配规则或者关键视频对象选择模型转化为由句法规则构成的表达式集合)进行关键视频对象选择可以看成是一次迭代,该迭代规则 $f(x)\rightarrow y$ 规定,满足句法规则的关键视频对象才是关键视频对象,故满足归纳法的基于最小不动点的迭代规则.
2. 可以根据经验设置度量函数的权值和阈值等,另外,可以指定第 1 个视频对象为关键视频对象等,这些都可以看成迭代的初始值,即作为归纳法的初始条件.
3. 只有满足迭代规则的视频对象才能加入关键视频对象集合,即满足归纳法的最小化条件.

故关键视频对象选择模型满足归纳法准则.

然后证明交互式关键视频对象选择模型满足余归纳法准则.

命题 2. 交互式关键视频对象选择模型满足余归纳法准则.

证明:

1. 开始时,由于用户并不知道确切的关键对象数,即没有用户交互时,所有的视频对象集合可以看成关键视频对象集合,故满足余归纳法的最大化条件.
2. 每次交互时,由于用户选择的策略不同,得到的结果也不同,满足定义 3 的交互条件.
3. 交互式关键视频对象选择模型满足余代数(定义 2),迭代规则 $\alpha_S: O \rightarrow O^S$ 通过交互进行动态绑定“全面”和“紧凑”策略,该迭代规则满足关键视频对象选择的最大可能性,则存在唯一的同态映射 $\alpha_S: O \rightarrow O^S$ (O^S 为偶对 (S, α) , α 为映射 $O \rightarrow S$, S 是一个状态集(由用户选择这两种策略)),同时,该同态映射满足余归纳法的基于最大不动点的迭代规则.

故交互式关键视频对象选择模型满足余归纳法准则.

由于交互式关键视频对象选择模型动态绑定“全面”和“紧凑”策略,因此能够动态产生关键视频对象集合,所以是一个动态系统.而关键视频对象选择模型产生的关键视频对象只能是静态关键视频对象集合.所以,从关键视频对象选择模型到交互式关键视频对象选择模型的转变,体现了从封闭系统到开放系统的扩展.

2 实验结果与分析

2.1 关键视频对象选择模型的实现

视频对象特征应该反映视频运动信息的显著变化,比如帧对象的颜色、纹理和形状等.本文的关键视频对象选择算法中主要利用了形状信息,而利用形状信息进行关键视频对象选择有以下两个优势:

1. 由于视频对象的运动,其结构相应地发生变化,如关节点变化等,可以通过对象形状的变化检测出来,而颜色和纹理信息有可能在整个运动中基本上保持不变.
2. 相对于纹理信息,形状信息更容易从帧中抽取,另外,颜色信息对某些领域如体育视频的区分性不好.

由于不变矩形状特征具有位移、旋转和缩放不变的良好性质,因此,本文的特征选取使用了改进的不变矩算子^[10],即 7 个不变矩 $\phi_i(i=1,2,\dots,7)$.为了提高区分度,对较小的 $\phi_i(i=3,4,\dots,7)$ 做了处理: $\phi'_i = \phi_i^{2/5}$.

为了方便比较,本文实现了 Kim^[4]中的基于不变矩形状特征的关键视频对象选择算法,其算法思想为:

1. 用分割工具将视频序列的背景和前景分离出来.
2. 提取前景中的视频对象.
3. 将视频对象用不变矩信息描述,即计算出加权的 7 个不变矩向量 F^0 .

- 第 1 帧的视频对象缺省作为第 1 个关键视频对象,然后依次比较相继帧的视频对象,若发现一个对象与当前关键视频对象在形状信息上变化显著,则该对象为新的关键视频对象,该帧为关键帧.如此反复直至结束,并显示关键帧的视频对象集合.

本文的实验数据为:用分割工具 AMOS^[11]提取 MPEG-4 的测试序列(bream 和 kettle)以及一些体育视频,包括跳水(A1,A2,A3,A4,A5,A6)、跳远(B)和排球(C).

2.2 交互式关键视频对象选择模型评估与分析

本文的交互式关键视频对象选择模型主要是对交互行为进行建模,为了给用户提供一个友好的界面,让用户根据自己的偏好选择关键视频对象选择.本文在界面上实现了关键视频对象的“全面”和“紧凑”的定性选择,分别表示基于最短路径约束的最小化失真率方法和基于最小化失真率约束的最短路径方法,其交互界面如图 1 所示.



Fig.1 Interactive key video objects selection interface

图 1 关键视频对象选择交互界面

由于跳水视频中跳水场景比较复杂,同时基于内容的跳水视频分析也是“数字奥运”中的重点研究课题,为此,本文主要针对跳水视频中的视频对象进行实验.以下为跳远视频的关键视频对象的实验结果.为了方便比较,根据“全面”和“紧凑”策略选择跳远和跳水视频的关键视频对象的实验结果分别如图 2(a)、图 2(b)所示, Kim^[4]方法的结果如图 2(c)所示.

由图 2 可以看出,本文提出的根据不同偏好(“全面”和“紧凑”)得出的关键视频对象均能够较好地反映相应体育运动的动作信息,其中“全面”策略得到的关键视频对象集合反映了体育视频中的主要信息,“紧凑”策略得到的关键视频对象集合简洁明了.

由于关键对象选择可以看成是数据压缩,因此不可避免地有信息损失,本文用失真率来评价关键对象选择的效果,其中,失真率 $D(O) = \sum_{i=0}^{n-1} d(o_i, o'_i) / n$, $\{o_i\}$ 是视频对象集合, $i=0, \dots, n-1$, $\{o'_i\}$ 是将关键对象帧替换原视频帧得到的集合, n 为视频对象总个数, $d(o_i, o'_i)$ 为 o_i 与 o'_i 的距离函数,在实验中设为帧间灰度图的像素差.

从表 1 可以看出,紧凑和全面策略的失真率略低于 Kim^[4]的方法,因此本文的方法可以产生满意的关键对象集合.

本文提出的方法实现简单,交互界面简便,而且从图 2 和图 3 来看,本文方法与 Kim^[4]方法相比有两个优势:

- 本文紧凑和全面策略的失真率低于 Kim^[4]的方法.
- Kim^[4]的方法是一个封闭系统,算法得到的关键视频对象集合是确定的,而本文的方法是一个开放系统,可以根据“全面”和“紧凑”策略反映视频的主要内容.

综上所述,交互式关键视频对象选择模型不仅能够动态地产生关键视频对象集合,而且这些集合能够反映视频对象运动的梗概.因此本文的交互式模型体现了从算法到交互的转变,不失为一种新的计算模型.

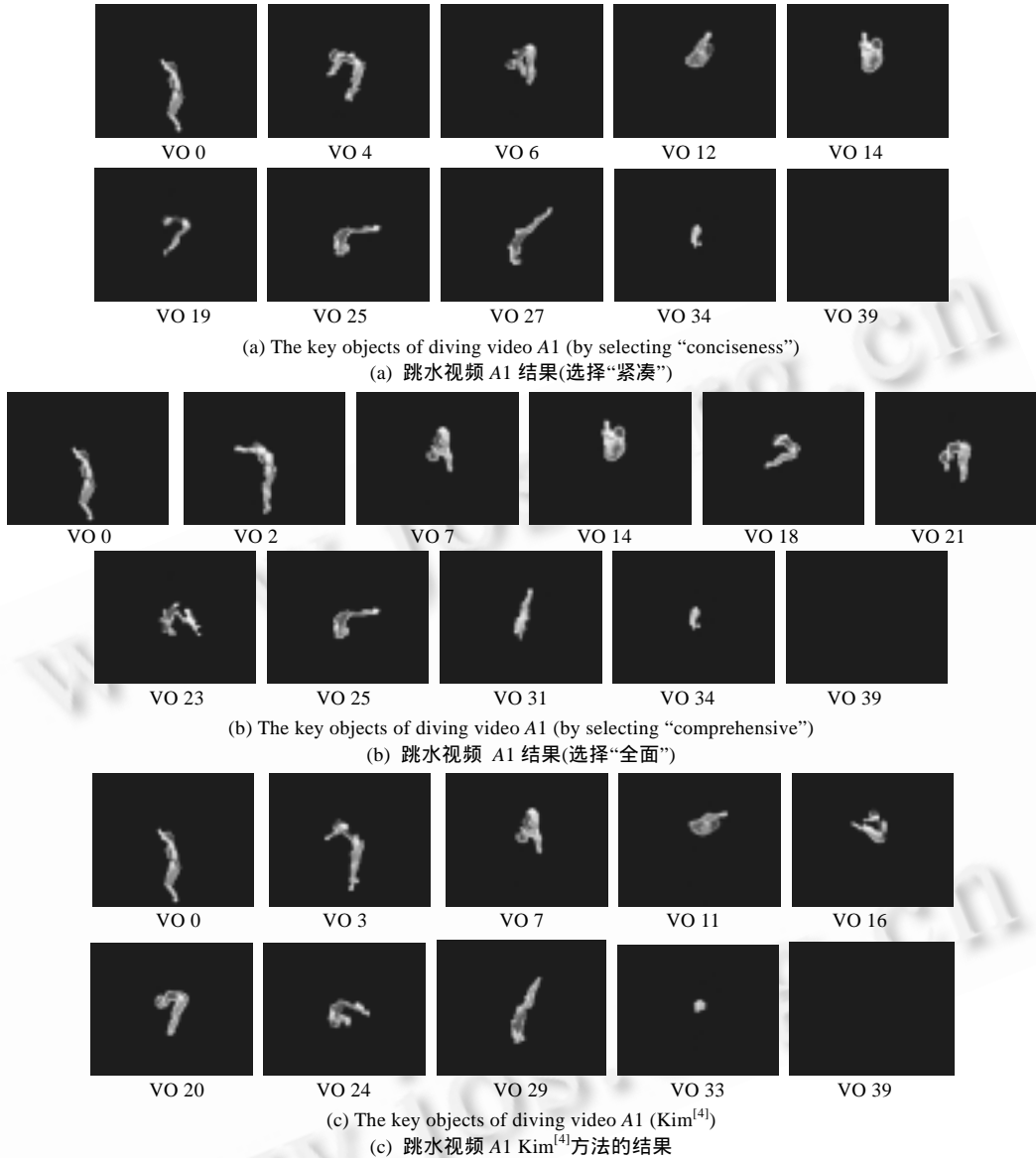


Fig.2 Results of "diving video A1" by "conciseness", "comprehensive" and Kim^[4]

图 2 跳水视频 A1“紧凑”、“全面”和 Kim^[4]的结果

Table 1 The distortion rate of selecting key video objects compared to the scheme of Kim^[4]

表 1 关键视频对象的失真率及与 Kim^[4]的比较

Video sequence	Frame number	"Conciseness"	"Comprehensive"	Kim ^[4]
Bream	198	0.04	0.04	0.04
Kettle	250	0.085	0.085	0.085
Diving video A1	40	0.009	0.008	0.009
Diving video A2	160	0.013	0.012	0.013
Diving video A3	29	0.009	0.009	0.01
Diving video A4	60	0.017	0.016	0.017
Diving video A5	95	0.024	0.023	0.024
Diving video A6	60	0.015	0.015	0.016
Long jump video B	125	0.005	0.003	0.005
Volley-Ball video C	56	0.02	0.019	0.02

3 结束语

当前的关键对象选取方法有一定的主观性,不能根据用户偏好动态地产生基于对象的视频摘要,因此是一个封闭系统.为此,本文提出一种基于交互计算的交互式关键视频对象选择模型,证明了关键视频对象选择模型满足归纳法准则,交互式关键视频对象选择模型满足余归纳法准则.实验表明,本文的紧凑和全面策略的失真率低于 Kim^[4]的方法.同时, Kim^[4]的方法是一个封闭系统,算法得到的关键视频对象集合是确定的,而本文的方法是一个开放系统,能够反映视频的主要内容.因此,交互式关键视频对象选择模型是对关键视频对象选择模型的扩展.

今后我们将对基于领域知识的经验进一步模型化,对用户交互行为进行建模,完善基于交互计算的关键视频对象选择模型的理论框架,并应用于中国的“数字奥运”工程.

References:

- [1] Zhang HJ, Wang JYA, Altunbasak Y. Content-Based video retrieval and compression: A unified solution. In: Proc. of the Int'l Conf. on Image Processing. 1997. 13–16. <http://citeseer.ist.psu.edu/zhang97contentbased.html>
- [2] Ferman AM, Gunsel B, Tekalp AM. Objects based indexing of MPEG-4 compressed video. In: Proc. of the SPIE Video Communication and Image Processing. vol. 3024. 1997. 953–963. <http://citeseer.ist.psu.edu/191203.html>
- [3] Erol B, Kossentini F. Automatic key video object plane selection using the shape information in the MPEG-4 compressed domain. IEEE Trans. on Multimedia, 2000,2(2):129–138.
- [4] Kim C, Hwang JN. Object-Based video abstraction for video surveillance systems. IEEE Trans. on Circuits and Systems for Video Technology, 2002,12(12):1128–1138.
- [5] Song X, Fan G. Key-Frame extraction for object-based video segmentation. In: Proc. of the 2005 IEEE Int'l Conf. on Acoustics, Speech, and Signal Processing. 2005. 689–692. <http://www.vcipl.okstate.edu/publications.html>
- [6] Wegner P. Tutorial notes: Models and paradigms of interaction. Technical Report, CS-95-21, Brown University, 1995. 1–67. <http://citeseer.ifi.unizh.ch/wegner95tutorial.html>
- [7] Wegner P, Goldin D. Mathematical models of interactive computing. Technical Report, CS 99-13, Brown University, 1999. 1–44. <http://www.cs.brown.edu/people/pw/>
- [8] Rutte JJMM. Universal coalgebra: A theory of systems. Technical Report, CS-R9652. Amsterdam: CWI (Centre for Mathematics and Computer Science), 1996. 1–55.
- [9] Nilsson NJ Write; Zheng KG, Zhuang YT, Trans. Artificial Intelligence: A new Synthesis.. Beijing: China Machine Press, 2000. 86–98 (in Chinese).
- [10] Flusser J. Fast calculation of geometric moments of binary images. In: Proc. of the 22nd Workshop Pattern Recognition Medical Computer Vision. 1998. 265–274. <http://citeseer.ist.psu.edu/flusser98fast.html>
- [11] Zhong D, Chang S. Amos: An active system for MPEG-4 video object segmentation. In: Proc. of the IEEE Int'l Conf. on Image Processing. 1998. 647–651. <http://www.ee.columbia.edu/dvmm/downloads/AMOSDownload.htm>

附中文参考文献:

- [9] Nilsson NJ, 著;郑扣根,庄越挺,译. Artificial Intelligence: A New Synthesis. 北京:机械工业出版社,2000.86–98.



欧阳建权(1973 -),男,湖南隆回人,博士,主要研究领域为人工智能,多媒体分析与检索,多值逻辑.



张勇东(1973 -),男,博士,CCF 高级会员,主要研究领域为视频编码与检索.



李锦涛(1962 -),男,博士,研究员,博士生导师,CCF 高级会员,主要研究领域为数字化技术.