

基于超图模型的复杂视频事件检测*

柯佳^{1a,1b}, 詹永照^{1b}, 陈潇君², 汪满容^{1c}

(1. 江苏大学 a. 管理学院; b. 计算机学院; c. 图书馆, 江苏镇江 212013; 2. 江苏大学附属医院信息科, 江苏镇江 212013)

摘要: 近年来,语义事件分析越来越受到重视,典型语义事件的检测与识别是一个具有挑战性的研究领域。提出了基于超图模型的复杂视频事件检测方法,通过分析对象的运动轨迹,检测出视频中的所有子事件并构建时序关系图及依赖关系图,从而生成子事件超图,并通过谱超图聚类分析来检测相应的复杂事件。采用图变换工具 AGG 进行模拟实验,其实验结果表明,该方法具有较高的准确率与召回率。

关键词: 视频语义; 视频事件检测; 时序关系; 超图模型; 谱超图聚类

中图分类号: TP391 **文献标志码:** A **文章编号:** 1001-3695(2012)12-4770-05

doi:10.3969/j.issn.1001-3695.2012.12.097

Detection of complexity video event based on hypergraph model

KE Jia^{1a,1b}, ZHAN Yong-zhao^{1b}, CHEN Xiao-jun², WANG Man-rong^{1c}

(1. a. School of Management, b. School of Computer Science & Telecommunication Engineering, c. Library, Jiangsu University, Zhenjiang Jiangsu 212013, China; 2. Information of Affiliated Hospital of Jiangsu University, Zhenjiang Jiangsu 212013, China)

Abstract: In recent years, semantic event analysis is more and more attention. To researchers, semantic event detection and recognition is a challenging area. This paper presented a method for detection of complex video event based on hypergraph model, which could analyze the trajectory of the object for detecting sub-events, built a timing diagram and dependency graph for generating a sub-events hypergraph, and used spectral hypergraph cluster for analyzing the complex event. It used the graph transformation tool AGG simulation to verify the new method. Experimental results show that this method has higher accuracy and recall rate.

Key words: video semantic; video event detection; temporal relationship; hypergraph model; spectral hypergraph clustering

0 引言

视频语义事件是指视频序列中具有一定语义信息的某种行为或者是一系列行为,而行为是指视频序列中一系列具有一定语义和时间延续性的动作。所谓事件检测就是指在视频中发现感兴趣的或者明显与众不同的事件,并对其包含的行为作出适当的分析。近年来,语义事件分析越来越受到重视,典型语义事件的检测与识别是一个具有挑战性的研究领域。针对语义事件的检测与识别,许多学者提出了各自不同的方法,主要分为以下三大类^[1]。

1) 基于预定义事件模型的事件检测方法 较早提出的一些事件检测方法使用预定义规则或限制条件来构建事件模型,如通过状态机来对事件建模。Badler^[2] 提出使用状态图和模拟环境下的原始规则构建事件模型的方法,该方法使用环境先验知识解决复杂的事件检测问题,而不是使用运动信息。Ayers 等人^[3] 也提出使用状态机来进行事件检测,但他们利用运动数据来进行事件检测,由于需要相关环境的先验知识,该方法只能用于特定领域下的事件检测,因而具有很大的局限性。

2) 基于训练数据学习事件模型的事件检测方法 基于训

练数据自动学习事件模型的事件检测方法^[4] 目前也有很多。通常这类方法在提取特征之后,采用隐马尔可夫模型(hidden markov model, HMM) 或者动态贝叶斯网络(dynamic Bayesian network, DBN) 等方法来分析各个关键帧特征值之间的关系,进而挖掘各个镜头之间的语义关系并检测出一些典型的事件。例如, Xin 等人^[5] 提出的对动态场景中语义事件检测方法;刘卫宁等人^[6] 提出的基于模糊 C-均值算法对监控视频中交通拥挤判断的研究。

3) 基于聚类分析的事件检测方法 该方法包括时空衍生^[7] 和协同嵌入式原型^[8] 等。上述两种方法都是通过对权重矩阵进行谱图分割来检测出事件片断,权重矩阵通过计算视频片断之间的相似性来确定。但是上述方法对事件的最大长度作了假定,并且仅局限于单主体事件以及单线程事件的检测。Rao 等人^[9] 提出了基于二维运动轨迹曲率分析的人体动作识别方法,该方法基于时空曲率分析得到相似事件的聚类,它没有对事件的长度作出假设,所以仅局限于单个主体事件的检测。

为改进和进一步发展事件检测方法,本文提出基于超图模型的视频复杂事件检测。首先,检测出视频中的运动对象并对其分类和跟踪,从而提取出对象的运动轨迹;然后,通过分析运

收稿日期: 2012-05-12; **修回日期:** 2012-06-24 **基金项目:** 国家自然科学基金资助项目(61170126); 国家教育部人文社会科学研究青年基金资助项目(12YJCZH020); 江苏省2008年度普通高校研究生科研创新计划资助项目(CX08B_097Z)

作者简介: 柯佳(1981-),女,江苏镇江人,讲师,博士研究生,主要研究方向为多媒体技术(kejia@ujs.edu.cn); 詹永照(1962-),男,教授,博士,主要研究方向为多媒体技术、模式识别; 陈潇君(1981-),男,工程师,硕士,主要研究方向为多媒体技术; 汪满容(1986-),男,硕士,主要研究方向为多媒体技术。

动轨,来迹检测出视频中的所有子事件并构建子事件时序关系图;再次,通过自动构建训练视频中各类子事件间的依赖关系图来建立复杂事件的模型;最后,依次利用每一个子事件依赖关系图来构建子事件超图,并通过谱超图聚类分析来检测相应的复杂事件。

1 基于超图模型的复杂事件建模

一个复杂事件实例通常可以划分成一系列具有一定时序关系的动作,定义为子事件。为了检测出一个视频片段中包含的所有子事件,首先需要检测出视频中的运动对象,并对其进行分类、跟踪,以便获取对象的运动轨迹。本文使用了本研究成果——改进的混合高斯模型方法^[10]来检测运动对象,采用标准化分割算法(normalized cuts, Ncuts)^[11]来分割对象的运动轨迹,从而划分出子事件。下面具体论述如何运用超图模型来实现复杂事件的建模过程。

1.1 超图理论

1970年 Berge^[12]首次提出了超图,并对超图理论进行了系统的阐述。超图的理论基础是图论和集合论。具有共同属性特征的对象属于一个集合,不同的抽象层次可归属于集合的集合,如此构成以集合的包含关系为基础的结构,这种结构可用超图来表示。超图的相关定义如下:

定义 1 设 $V = \{v_1, v_2, \dots, v_n\}$ 是 n 个顶点的有限集。 V 的某一子集被称为 e_i , 若 $\bigcup_{i=1}^m e_i = V$, 则称 $E = \{e_1, e_2, \dots, e_m\}$ 为超边集, 称 $H = (V, E)$ 为超图。超图 H 可由一个 $|V| \times |E|$ 的矩阵 A 表示, 即

$$a(v_i, e_j) = \begin{cases} 1 & v_i \in e_j \\ 0 & v_i \notin e_j \end{cases} \quad (1)$$

称 A 为超图 H 的关联矩阵。

若在超图的超边集中定义了方向, 则超图为有向超图。有向超图是超图概念的扩展。其定义如下:

定义 2 设 $H = (V, E)$ 为超图, 若 $E = \{e_1, e_2, \dots, e_m\}$ 为有向超边集, 则称 H 为有向超图。

定义 3 若在超边集 E 上定义权函数, 并且设 E 的权值集为 $W = \{w_1, w_2, \dots, w_m\}$, 则超图 $H = (V, E, W)$ 称为加权超图。

定义 4 顶点 $v_i \in V$ 的度 $d(v_i) = \sum_{\{e_j \in E | v_i \in e_j\}} w(e_j) = \sum_{e_j \in E} w(e_j) h(v_i, e_j)$, 其中 $w(e_j)$ 表示 e_j 对应的权值; 超边 $e_j \in E$ 的度 $\delta(e_j) = |e_j| = \sum_{v_i \in V} h(v_i, e_j)$ 。

有关超图的表示分为无向超图和有向超图两种情况。在无向超图中, 若 $|e_i| > 2$, 则画成一曲线包围 e_i 中的所有顶点; 若 $|e_i| = 2$, 则画成一曲线连接两顶点; 若 $|e_i| = 1$ 则为自环。有向超图的一种表示法是超边、端点和引线表示法^[13], 即在 e_i 的每一端引出一根线, 称为引线。若 $e_i \cap e_j \neq \emptyset$, 则在其交集的端点用引线连接。

超图的应用领域比较广泛, 基于超图的典型聚类算法是将高维空间中的原始数据及其相互关系映射到带权超图 $H = (V, E)$ 。其中: V 表示数据项点集, E 表示连接相关数据项子集的超边集。设 w_i 为对应于 E 中每一条超边 $e_i \in E$ 的权重, 用于衡量超边连接的多个相关数据项之间的相关程度。同时给出评价函数 $f = \sum w_i$ 。在数据集的超图确定后, 选用一种图形分割最优算法对超图进行分割。分割依据是令评价函数 $f = \sum w_i$ 的值最小, 即令各子分割之间相关性之和最小。

1.2 子事件时序关系图

视频中子事件的发生时间顺序各不相同, 子事件时序关系图(sub-event temporal relationship graph, STRG)就是为了描述各子事件之间的时序关系。为了能够更加具体地描述子事件之间的时序关系, 引入了文献[14]中提出的时序逻辑, 可以将子事件之间的时序关系划分为七种类型, 即 After、Meet、Overlaps、Finishes、Contain、Starts、Equal。设 SE_1 与 SE_2 均为视频子事件, 则它们之间的时序关系可用图 1 来说明。具体来说, 设 SE_i^{Begin} 与 SE_i^{End} 分别表示子事件 SE_i ($i = 1, 2$) 的开始时间和结束时间, 则子事件 SE_1 与 SE_2 之间的各种时序关系可定义如下:

$$\begin{aligned} \text{After: } & SE_2^{\text{Begin}} > SE_1^{\text{End}} \\ \text{Meet: } & SE_2^{\text{Begin}} = SE_1^{\text{End}} \\ \text{Overlaps: } & SE_2^{\text{Begin}} > SE_1^{\text{Begin}} \quad \text{and} \quad SE_2^{\text{Begin}} < SE_1^{\text{End}} \quad \text{and} \quad SE_2^{\text{End}} > SE_1^{\text{End}} \\ \text{Finishes: } & SE_2^{\text{Begin}} > SE_1^{\text{Begin}} \quad \text{and} \quad SE_2^{\text{End}} = SE_1^{\text{End}} \\ \text{Contain: } & SE_2^{\text{Begin}} > SE_1^{\text{Begin}} \quad \text{and} \quad SE_2^{\text{End}} < SE_1^{\text{End}} \\ \text{Starts: } & SE_2^{\text{Begin}} = SE_1^{\text{Begin}} \quad \text{and} \quad SE_2^{\text{End}} > SE_1^{\text{End}} \\ \text{Equal: } & SE_2^{\text{Begin}} = SE_1^{\text{Begin}} \quad \text{and} \quad SE_2^{\text{End}} = SE_1^{\text{End}} \end{aligned}$$

基于上述时序关系, 视频中各子事件之间的时序关系可以用一个有向无环图(directed acyclic graph, DAG)来表示, 图中的每个顶点代表视频中的一个子事件, 而其中的每条有向边则代表分别与该边的起点和终点对应的两个子事件之间的时序关系。将一个 STRG 形式化地表示为 $G = (V, E)$, 其中顶点集 $V = \{v_1, v_2, \dots, v_n\}$, 有向边集 $E = \{e_1, e_2, \dots, e_m\}$ 。对于一个给定的视频, 设检测得到的子事件集为 $SESet = \{SE_1, SE_2, \dots, SE_n\}$, 则该视频的子事件时序关系图构建过程如下:

a) 对 $SESet$ 中的各子事件按照开始时间升序排列(开始时间相同时按结束时间升序排列), 排序之后得到的子事件序列 $OrderedSESet = \{OSE_1, OSE_2, \dots, OSE_n\}$ 。

b) 依次为 $OrderedSESet$ 中的每个子事件建立一个顶点, 得到相应的有序顶点序列 $V = \{v_1, v_2, \dots, v_n\}$ 。

c) 对于顶点集 V 中的每个顶点 v_i ($i = 1, 2, \dots, n-1$), 依次构造由顶点 v_i 指向其后续顶点 v_j ($j = i+1, i+2, \dots, n$) 的有向边, 最终可得到有向边集 $E = \{e_1, e_2, \dots, e_m\}$ 。

对于 STRG 中的每条有向边 e_i ($i = 1, 2, \dots, m$), 均为其分配一个时序关系标签, 用于描述分别与该边的起点和终点对应的两个子事件之间的时序关系。因此, 一个 STRG 的有向边集 $E = \{e_1, e_2, \dots, e_m\}$ 对应一个时序关系标签集 $TR = \{tr_1, tr_2, \dots, tr_m\}$, 其中 $tr_k \in TRTypes = \{\text{After, Meet, Overlaps, Finishes, Contain, Starts, Equal}\}$ ($k = 1, 2, \dots, m$)。为了便于更加直观地理解子事件时序关系图, 图 2 给出了一个在测试视频中检测到的船只进港并靠近油轮事件所对应的子事件时序关系。

1.3 子事件依赖关系图

对于一个特定复杂事件而言, 其包含的各个子事件之间通常存在不同程度的依赖关系。子事件依赖关系图(sub-event dependency graph, SDG)就是用来描述一个事件所包含的各个子事件之间的时序条件依赖关系, 它可以有效地表示事件的结构。给定一个特定事件的训练视频, 通过自动构建训练视频中各类子事件间的依赖关系图来建立该事件的模型。

在同一个事件中, 一个子事件的发生通常与跟它邻近的若干前驱子事件的发生相关, 也就是说, 子事件之间的依赖关系

属于高阶马尔可夫依赖。因此,为了有效地描述子事件间的高阶马尔可夫依赖关系,采用加权有向超图(weighted directed hypergraph,WDH)来表示子事件依赖关系图。形式化地,将SDG表示为一个超图 $G = (V, E, W)$,其中顶点集 $V = \{v_1, v_2, \dots, v_n\}$,超弧集 $E = \{e_1, e_2, \dots, e_m\}$,超弧的权值集 $W = \{w_1, w_2, \dots, w_m\}$ 。

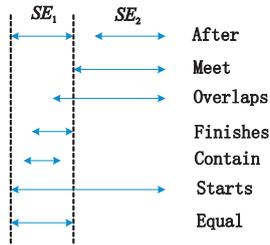


图1 子事件 SE_1 与 SE_2 之间的时序关系

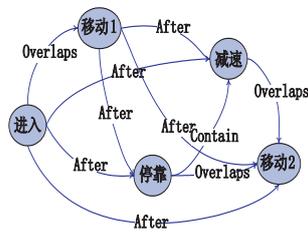


图2 船只进港并靠近油轮事件对应的子事件时序关系

在一个SDG中,每个顶点表示可能在一个事件中发生的某个子事件,而每条超弧则表示该超弧包含的顶点所对应的子事件之间存在时序条件依赖关系。设一个子事件最多与跟它邻近的 x 个前驱子事件相关,则在SDG中任意一条超弧最多具有 $x + 1$ 个顶点。设超弧 $e_k = \langle P_i, v_i \rangle$,则有序顶点集 $P_i \subseteq V$,且 P_i 中的各顶点所对应的子事件在时序上邻近顶点 v_i 对应的子事件。超弧 e_k 对应的权值 w_k 表示顶点 v_i 对应的子事件与 P_i 中的各顶点所对应的子事件之间的关联程度,具体来说就是指在 P_i 中的各顶点所对应的子事件依次发生的条件下顶点 v_i 对应的子事件发生的概率。由此可知,当超弧 e_k 包含的顶点个数 $X \in \{2, 3, 4\}$ 时,按照式(2)~(4)来计算其相应的权值 w_k :

$$P(SE_i^t | SE_j^{t-1}) = \frac{P(SE_i^t, SE_j^{t-1})}{P(SE_j^{t-1})} \quad (2)$$

$$P(SE_i^t | SE_j^{t-1}, SE_k^{t-2}) = \frac{P(SE_i^t, SE_j^{t-1}, SE_k^{t-2})}{P(SE_j^{t-1} | SE_k^{t-2}) P(SE_k^{t-2})} \quad (3)$$

$$P(SE_i^t | SE_j^{t-1}, SE_k^{t-2}, SE_l^{t-3}) = \frac{P(SE_i^t, SE_j^{t-1}, SE_k^{t-2}, SE_l^{t-3})}{P(SE_j^{t-1} | SE_k^{t-2}, SE_l^{t-3}) P(SE_k^{t-2} | SE_l^{t-3}) P(SE_l^{t-3})} \quad (4)$$

其中: SE_i^t 为顶点 v_i 对应的子事件, SE_i^{t-3} 、 SE_k^{t-2} 、 SE_j^{t-1} 依次为有序顶点集 P_i 中的各顶点所对应的子事件。

对于一个特定复杂事件而言,采用训练视频自动构建该事件的SDG,从而为以后的事件检测奠定基础。基于训练视频的子事件时序关系图,获取一个时序上有序的子事件序列 $OrderedSESeq = \{SE_1, SE_2, \dots, SE_n\}$ 。利用训练视频的子事件序列 $OrderedSESeq$,按照以下步骤自动构建事件的SDG:

a) 依次判断 $OrderedSESeq$ 中的每个子事件 $SE_i (i = 1, 2, \dots, n)$ 所属的类别 $C(SE_i)$,汇总得到相应的子事件类别集 $SETypes = \{c_1, c_2, \dots, c_m\}$ 。

b) 令子事件类别序列集 $SECSeqSet = \emptyset$,且设 $AG(\cdot)$ 表示相应子事件的主体集,按先后顺序对 $OrderedSESeq$ 中的每个子事件 $SE_i (i = 1, 2, \dots, n)$ 执行以下操作:

(a) 在 SE_i 之后寻找与之最邻近的三个子事件 $\{SE_j, SE_k, SE_l\}$,其中 $i < j < k < l \leq n$ 且 $AG(SE_i) \cap AG(SE_j) \neq \emptyset$ and $AG(SE_j) \cap AG(SE_k) \neq \emptyset$ and $AG(SE_k) \cap AG(SE_l) \neq \emptyset$;

(b) 对 $\{SE_j, SE_k, SE_l\}$ 中的每个子事件 $SE_p (p = j, k, l)$ (如果存在),若类别序列 $C(SE_i)C(SE_j) \dots C(SE_p) \in SECSeqSet$,则将该类别序列出现的频数加1,否则将该类别序列出现的频

数设为1,并令 $SECSeqSet = SECSeqSet \cup \{C(SE_i)C(SE_j) \dots C(SE_p)\}$ 。

c) 依次为 $SETypes$ 中的每个子事件类别建立一个顶点,得到SDG的顶点集 $V = \{v_1, v_2, \dots, v_m\}$ 。

d) 对于 $SECSeqSet$ 中的每个类别序列 $c_i c_j \dots c_p (p = j, k, l; 1 \leq i, j, k, l \leq m)$,为SDG构造一条相应的超弧 $e_r = \langle v_i, v_j, \dots, v_p \rangle$,并且 e_r 的权值 $w_r = \frac{Freq(c_i c_j \dots c_p)}{\sum_{c_q \in SETypes} Freq(c_i c_j \dots c_q)}$,其中 $Freq(\cdot)$

表示相应的类别序列出现的频数。

通过以上分析可知,采用SDG来对复杂事件建模,不仅能很好地捕捉同一种事件中各个子事件之间可能出现的时序关系变化,而且能有效地描述各个子事件之间不同大小的时序条件依赖关系。更加值得注意的是,SDG还可以用于多主体事件的建模,且用它来对事件建模时对事件的长度没有限制。

2 基于谱超图聚类的复杂事件检测

对于一个给定的测试视频,首先检测得到视频中的所有子事件并构建相应的子事件时序关系图;其次采用每一个已经构建好的子事件依赖关系图来构建测试视频的子事件超图;最后通过谱超图分割来获取所有关联度大的子事件簇,其中的每一个子事件簇均组成了一个相应的子事件依赖关系图所对应的复杂事件。因而,只要依次利用每一个子事件依赖关系图来构建测试视频的子事件超图并对其进行谱超图聚类分析,就可以检测出测试视频中的所有事件。

2.1 构建子事件超图

基于测试视频的子事件时序关系图,可以获取一个时序上有序的子事件序列 $OrderedSESeq = \{SE_1, SE_2, \dots, SE_n\}$ 。给定一个特定复杂事件的子事件依赖关系图 SDG_s ,按照以下步骤构建测试视频的子事件超图:

a) 依次为 $OrderedSESeq$ 中的每个子事件建立一个顶点,得到子事件超图的顶点集 $V = \{v_1, v_2, \dots, v_n\}$ 。

b) 设 $AG(\cdot)$ 表示相应子事件的主体集, $C(\cdot)$ 表示相应子事件的类别,按先后顺序对 $OrderedSESeq$ 中的每个子事件 $SE_i (i = 1, 2, \dots, n)$ 执行以下操作:

(a) 在 SE_i 之后寻找与之最邻近的三个子事件 $\{SE_j, SE_k, SE_l\}$,其中 $i < j < k < l \leq n$ 且 $AG(SE_i) \cap AG(SE_j) \neq \emptyset$ and $AG(SE_j) \cap AG(SE_k) \neq \emptyset$ and $AG(SE_k) \cap AG(SE_l) \neq \emptyset$;

(b) 对 $\{SE_j, SE_k, SE_l\}$ 中的每个子事件 $SE_p (p = j, k, l)$ (如果存在),若类别序列 $C(SE_i)C(SE_j) \dots C(SE_p)$ 在 SDG_s 中有一条超弧与之对应,则为子事件超图中的顶点 $\{v_i, v_j, \dots, v_p\}$ 构造一条超边,该超边的权值 w_a 由以下公式计算:

$$w_a = P(SE_p | Pa(SE_p)) \quad (5)$$

其中: $Pa(SE_p)$ 表示 SE_p 的前驱子事件序列 SE_i, SE_j, \dots 。

通过以上描述可知,那些在训练视频中频繁出现的子事件序列在子事件超图中所对应的超边会具有较高的权值,因而这些子事件在后续的谱超图聚类中就会被划分在同一个类别中。

2.2 子事件超图的谱分割

谱超图分割是谱图分割在超图上的扩展,它可以有效地将一个超图分割成两部分。给定一个超图 $H = (V, E, W)$,设顶点集 $S \subseteq V, S^c$ 为 S 的补集,则谱分割通过实现以下条件来得到 H 的最优二分:

$$\operatorname{argmin}_{\emptyset \neq S \subseteq V} c(S) = \operatorname{vol} \partial S \left(\frac{1}{\operatorname{vol} S} + \frac{1}{\operatorname{vol} S^c} \right) \quad (6)$$

其中: $\partial S = \{e \in E \mid e \cap S \neq \emptyset, e \cap S^c \neq \emptyset\}$ 为顶点集 S 的超割边集, $\operatorname{vol} \partial S = \sum_{e \in \partial S} w(e) \frac{|e \cap S| |e \cap S^c|}{\delta(e)}$, $\operatorname{vol} S = \sum_{v \in S} d(v)$, $\operatorname{vol} S^c = \sum_{v \in S^c} d(v)$ 。也就是说,通过满足式(6)来分割超图,可以使分割得到的两部分各自内部关联最紧密而两部分之间的关联最小。

为了得到满足上述条件的分割,通过求解以下特征系统的最小非零特征值所对应的特征向量来解决:

$$(I - D_v^{-1/2} A W D_e^{-1} A^T D_v^{-1/2}) x = \lambda x \quad (7)$$

其中: I 为单位矩阵; A 为超图 H 的关联矩阵; A^T 为 A 的转置; D_e 是由各个顶点的度构成的对角阵; D_v 是由各条超边的度构成的对角阵; W 是由各条超边的权值构成的对角阵。若设 $\Delta = I - D_v^{-1/2} A W D_e^{-1} A^T D_v^{-1/2}$, 并设 Φ 为 Δ 的最小非零特征值对应的特征向量,则超图 H 的最优二分 $S = \{v \in V \mid \Phi(v) \geq 0\}$, $S^c = \{v \in V \mid \Phi(v) < 0\}$ 。

基于上述方法,对构建好的子事件超图递归地进行二分,直到 $c(S)$ 的值以及分割所得各部分内超边的平均权值均不小于相应的阈值。最终可以获取子事件超图中的所有子事件簇,并且同一个子事件簇中的各个子事件间的关联程度很大,而不同簇中的子事件之间的关联程度很小。

由于子事件超图是根据一个特定复杂事件的子事件依赖关系图构建的,因此只能确定每个子事件簇是否对应一个该特定复杂事件的实例。对于每一个子事件簇,若其中的各个子事件之间的时序依赖关系与相应的特定复杂事件中的各个子事件之间的时序依赖关系相似,则该子事件簇就对应一个相应的特定复杂事件的实例;反之不然。对于子事件簇 SEC_i 与特定的复杂事件 SDG_j 之间的相似度,按照以下公式计算:

$$\operatorname{Sim}(SEC_i, SDG_j) = \frac{VN_{in}}{VN_{SDC}} \times \frac{EN_{in}}{EN_{total}} \quad (8)$$

其中: VN_{in} 表示 SEC_i 中包含在 SDG_j 中的顶点的数量; VN_{SDC} 表示 SDG_j 中的顶点数量; EN_{in} 表示 SEC_i 中包含在 SDG_j 中的超边的数量; EN_{total} 表示 SEC_i 中包含在 SDG_j 中的各个顶点之间的超边的总数。

3 系统测试

为了验证上述方法的有效性,选择镇江海关监控系统中采集的监控视频集 (customs monitoring video, CMV) 为实验数据,并针对其中的 5 个典型事件进行了实验,各个事件的编号及其描述信息如表 1 所示。

为评估 SDG 方法的正确性和一致性,采用图变换 (graph transformation) 工具 AGG 进行实验。AGG 是一个基于代数方法的图变换系统仿真环境,它不仅可以通过图形方式模拟图变换系统的变换过程,而且还可检查图变换系统的一致性。港口码头 (pier) 中的运动对象油轮 (tanker)、小船 (vessel) 之间的事件关系如图 3(a) 所示,由此构成了以上 5 个 Event 的子事件时序关系图,实验的时序性按照视频帧过场顺序;由所有的视频集数据传入子事件时序关系图进行计算,得到各个子事件超图的构建过程如图 3(b) 所示,由于子事件超图就可以进行谱分割,从而得到事件监测模型,用于复杂视频事件的检测。

表 1 各事件的详细信息

事件编号	事件描述	总帧数	子事件数量	事件数量
1	油轮靠港,卸油,停止作业	12 870	36	5
2	船只靠近油轮,船只停靠油轮,开始输油,停止作业	62 271	108	15
3	桥起重机开始移动,移动至油轮上方,停止移动	15 477	42	7
4	船只靠近油轮,检查,船只停留油轮侧方	23 595	57	10
5	船只靠近油轮,船只经过油轮旁边区域	21 714	51	9

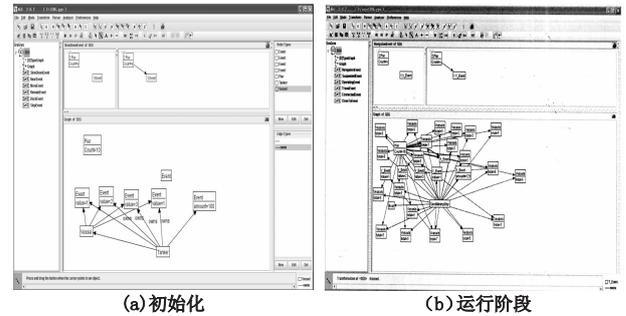


图 3 复杂视频事件的检测仿真图

为了客观地评价 SDG 方法的合理性和有效性,选取准确率与召回率两个指标作为衡量标准,下面给出几个相关术语。

真正 (true positive, TP), 定义为被分类模型正确预测的正样本数。

假正 (false positive, FP), 定义为被分类模型错误预测为正类的负样本数。

真负 (true negative, TN), 定义为被分类模型正确预测的负样本数。

假负 (false negative, FN), 定义为被分类模型错误预测为负类的正样本数。

准确率 precision 的计算公式为

$$\operatorname{precision} = \frac{TP}{TP + FP} \quad (9)$$

召回率 recall 的计算公式为

$$\operatorname{recall} = \frac{TP}{TP + FN} \quad (10)$$

通过上述定义可知,一种好的事件检测方法应该具有较高的准确率和较高的召回率。

在实验过程中,将 SDG 方法、基于谱图聚类的方法 (CGC)、基于时空衍生的方法 (MDL)^[7] 和协同嵌入式原型 (NDE)^[8] 进行了对比,四种方法的准确率和召回率及其对比如图 4 所示。

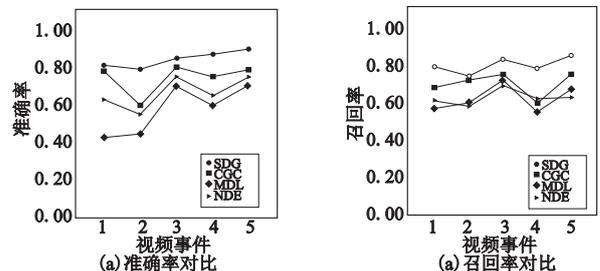


图 4 准确率及召回率对比

如图 4(a) 所示,SDG 方法的准确率和召回率均高于其他方法。对于事件 2,SDG 方法的准确率明显高于其他算法,这

是由于事件 2 比较复杂,其他算法很容易将一个与事件 2 相似的事件误判为一个事件 2 的实例(由于复杂事件比较容易被截断);对于事件 4,针对小船长时间停留情况,SDG 方法也可以得到很好的准确率。如图 4(b)所示,对于事件 5,本文方法的召回率比较高,原因是事件 5 时间间隔较短,SDG 方法可以更好地区分简短事件的子事件与其他干扰子事件。所以 SDG 方法用于事件检测能得到较好的准确率和召回率,能够有效地解决复杂视频事件的检测问题。

4 结束语

本文提出了一种基于超图模型的视频复杂事件建模及检测方法。该方法首先检测出视频中的运动对象并对其进行分类和跟踪,从而提取出对象的运动轨迹;然后通过分析对象的运动轨迹,检测出视频中的所有子事件并构建子事件时序关系图。基于子事件时序关系图,对于训练视频,通过自动构建训练视频中各类子事件间的依赖关系图来建立复杂事件的模型;而对于测试视频,依次利用每一个子事件依赖关系图来构建子事件超图,并通过谱超图聚类分析来检测相应的事件。采用 SDG 来对复杂事件建模不仅能很好地捕捉同一种事件中各个子事件之间可能出现的时序关系变化,而且能有效地描述各个子事件之间不同大小的时序条件依赖关系。不仅如此,SDG 还可以用于多主体事件的建模,且用它来对事件建模时对事件的长度没有限制。因而通过对子事件超图的谱聚类分析,不仅可以检测出多主体事件,而且还可以检测出任意长度的事件。实验结果表明,SDG 方法具有更高的准确率和召回率,能更加有效地适应复杂视频事件的检测。

参考文献:

- [1] HAKEEM A, SHAH M. Learning, detection and representation of multi-agent events in videos[J]. *Artificial Intelligence*, 2007, 171(8-9): 586-605.
- [2] BADLER N. Temporal scene analysis: conceptual description of object movements, University of Toronto Technical Report No. 80[R].

1975.

- [3] AYERS D, SHAH M. Monitoring human behavior from video taken in an office environment[J]. *Image and Vision Computing*, 2001, 19(12): 833-846.
- [4] 徐杨,吴成东,陈东岳. 基于视频图像的交通事件自动检测算法综述[J]. *计算机应用研究*, 2011, 28(4): 1206-1210.
- [5] XIN Lun, TAN Tie-niu. Semi-supervised learning on semantic manifold for event analysis in dynamic scenes[C]//Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2007: 18-23.
- [6] 刘卫宁,曾恒,孙隼华,等. 基于视频检测技术的交通拥挤判别模型[J]. *计算机应用研究*, 2010, 27(8): 3006-3008.
- [7] ZHANG Zhang, HUANG Kai-qi, TAN Tie-niu, *et al.* Trajectory series analysis based event rule induction for visual surveillance[C]//Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2007: 1-8.
- [8] ZHONG Hua, SHI Jian-bo, VISONTAI M. Detecting unusual activity in video[C]//Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2004: 819-826.
- [9] RAO Cen, YILMAZ A, SHAH M. View-invariant representation and recognition of actions[J]. *International Journal of Computer Vision*, 2002, 50(2): 203-226.
- [10] JU Shi-guang, CHEN Xiao-jun, XU Guang-hua. An improved mixture Gaussian models to detect moving object under real-time complex background[C]//Proc of International Conference on Cyberworlds. 2008: 730-734.
- [11] SHI Jian-bo, MALIK J. Normalized cuts and image segmentation[J]. *IEEE Trans of Pattern Analysis and Machine Intelligence*, 2000, 22(8): 888-905.
- [12] BERGE C. Graph and hypergraph[M]. Amsterdam: North-Holland, 1973: 528.
- [13] 黄汝激. 超网络的有向超树分析法[J]. *电子科学学刊*, 1987, 19(3): 244-255.
- [14] ALLEN J F, FERGUSON G. Actions and events in interval temporal logic[J]. *Journal of Logic Computation*, 1994, 14(5): 531-579.

(上接第 4769 页)

参考文献:

- [1] JI Shui-wang, YE Jie-ping. Generalized linear discriminant analysis: a unified framework and efficient model selection[J]. *IEEE Trans on Neural Networks*, 2008, 19(10): 1768-1782.
- [2] CHEN Li-fen, LIAO H Y M, KO M T, *et al.* A new LDA-based face recognition system which can solve the small sample size problem[J]. *Pattern Recognition*, 2000, 33(10): 1713-1726.
- [3] CHEN Song-can, LI Dao-hong. Modified linear discriminant analysis[J]. *Pattern Recognition*, 2005, 38(3): 441-443.
- [4] LI Hai-feng, JIANG Tao, ZHANG Ke-shu. Efficient and robust feature extraction by maximum margin criterion[C]//Advances in Neural Information Processing Systems. 2003: 157-165.
- [5] YANG Jian, ZHANG D, FRANCI A F. Two dimensional PCA: a new approach to appearance-based face representation and recognition[J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2004, 26(1): 131-137.

- [6] XIONG Hui-lin, SWAMY M N, AHMAD M O. Two-dimensional FLD for face recognition[J]. *Pattern Recognition*, 2005, 38(7): 1121-1124.
- [7] YANG Jian, ZHANG D, YONG Xu, *et al.* Two-dimensional discriminant transform for face recognition[J]. *Pattern Recognition*, 2005, 38(7): 1125-1129.
- [8] YE Jie-ping, JANARDAN R, LI Qi. Two-dimensional linear discriminant analysis[C]//Advances in Neural Information Processing Systems. 2004: 1569-1576.
- [9] HE Xiao-fei, CAI Deng, NIYOGE P. Tensor subspace analysis[C]//Advances in Neural Information Processing Systems. 2005.
- [10] GU Quan-quan, ZHONG Jie. Two dimensional maximum margin criterion[C]//Proc of IEEE International Conference on Acoustics, Speech and Signal Processing. 2009.
- [11] YAN Shui-cheng, XU Dong, ZHANG Lei, *et al.* Coupled kernel-based subspace learning[C]//Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2005: 645-650.