

基于时空语义信息的视频运动目标交互行为识别方法

金 标^{1,2,3} 胡文龙^{1,2} 王宏琦^{1,2}

¹ 中国科学院电子学研究所, 北京 100190
² 中国科学院空间信息处理与应用系统技术重点实验室, 北京 100190
³ 中国科学院研究生院, 北京 100049

摘要 提出一种融合时间及目标之间空间语义信息的视频运动目标交互行为识别方法,即基于目标之间空间语义的变化规律识别其交互行为类别。不同于传统的语义事件建模方法,首先根据运动目标跟踪结果,基于其运动方向以及建立目标之间的空间关系(拓扑关系和方向关系)模型,提出一种提取人目标之间空间语义(前面、后面、背对、面对以及左右)的方法;然后基于空间语义的变化规律建立随机文法规则;最后采用随机文法器识别九种常见的两人交互行为。该方法无需训练样本,实验结果验证了方法的有效性及其优越性。

关键词 机器视觉;交互行为识别;空间关系;空间语义;随机文法

中图分类号 TP391 **文献标识码** A **doi**: 10.3788/AOS201232.0515002

Moving-Objects Interaction Recognition Based on the Spatial-Temporal Semantic Information

Jin Biao^{1,2,3} Hu Wenlong^{1,2} Wang Hongqi^{1,2}

¹ Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China
² Key Laboratory of Spatial Information Processing and Application System Technology, Chinese Academy of Sciences, Beijing 100190, China
³ Graduate University of Chinese Academy of Sciences, Beijing 100049, China

Abstract A method for recognizing human-human interaction is proposed based on the spatial-temporal semantic information. Different from traditional methods to model the interactions, this framework achieves recognizing activities based on the transformation of the spatial semantic meaning. First, with detection and tracking results, the spatial semantic meaning (front, back, face to face, back to back, and left or right) between the persons are extracted based on motion directions and spatial relationships (including topological and directional relations). Then, stochastic context-free grammar is used to recognize interactions that the rules are learned based on the transformation of spatial semantics. Extensive experiments have been executed to validate the effectiveness of the proposed approach, and the method can recognize the interactions without additional training.

Key words machine vision; interaction recognition; spatial relationship; spatial semantic meaning; stochastic context-free grammar

OCIS codes 150.0155; 100.4999; 100.5010; 140.1135

1 引 言

视频运动目标交互行为分析在智能视频监控、视频注解、虚拟现实和人机交互等领域中具有广阔的应用前景,已经成为计算机视觉以及模式识别等

学科的研究热点。然而传统的交互行为识别方法需要大量样本训练模型,其性能不稳定,实用性也较差。

目前,视频运动目标行为识别方法^[1~3]主要包

收稿日期: 2011-12-07; **收到修改稿日期**: 2012-01-03

基金项目: 国家 973 计划(2010CB327900)和国家自然科学基金(61001176)资助课题。

作者简介: 金 标(1984—),男,博士研究生,主要从事计算机视觉和图像理解等方面的研究。

E-mail: loyierve@126.com

导师简介: 胡文龙(1963—),男,博士,研究员,主要从事信号与信息处理方面的研究。E-mail: hwen2k@gmail.com

括模板匹配、基于有限状态机的方法、基于概率图模型的方法以及基于随机文法的方法等。然而大部分方法是通过提取简单原子事件或轨迹序列等动态特征序列,然后分析动态特征序列识别行为,通常需要人为标定训练样本,之后进行模板的训练和模型参数的学习^[1,2]。采用模板匹配进行行为识别的方法^[4,5]首先将图像序列转换为一组静态形状模式,然后在识别过程中和预先存储的行为模板相比较,例如 Bobick^[4]采用运动能量图像(MEI)和运动历史图像(MHI)来解释图像系列中人的运动,模板匹配方法的优点是计算复杂度低,实现简单,然而其对于噪声和运动时间间隔的变化较敏感。有限状态机将特征序列表示为几个行为状态的转移,利用状态转移函数,确定输入特征序列所代表的最佳状态序列,并用该状态序列确定输入特征序列与参考序列是否匹配,每一个状态代表一个简单的原子事件^[6]。概率图模型是当前最为流行的为动态特征序列建立语义概念模型的工具^[7~13],其最大优点在于引入一个概率框架,在此框架下,该模型可对输入的特征序列进行概率判断解决观测值的不确定性,可分为以下几个常用的模型:贝叶斯网、隐马尔科夫模型和动态贝叶斯网等。随机文法识别方法将简单原子事件作为随机文法器的输入,进而识别行为^[14~16]。然而现有大部分方法需要学习并建模简单原子事件,其适应性较差,同时模型学习时间较长。

不同于传统方法,提出一种融合时间及目标之间空间语义信息的视频人目标交互行为识别方法,即基于目标之间空间语义的变化规律识别其交互行为类别,该方法无需训练样本。首先,基于运动目标

检测与跟踪结果,建立目标之间的空间关系(拓扑关系和方向关系)模型,提取目标之间的空间语义(前面、后面、背对、面对以及左右);根据实际经验,基于空间语义的变化规律建立随机文法规则;最后采用随机文法器识别常见的交互行为。

2 空间语义

提出的空间语义是指目标对象与参考对象之间的位置信息,主要包括前面、后面、背对、面对以及左右等,本节通过建模空间关系进而提取空间语义。

2.1 空间关系建模

空间关系主要包括拓扑关系、方向关系和度量关系。其中拓扑和方向关系属于定性关系,而度量关系属于定量关系。本文研究只涉及面目标与面目标之间的定性关系。

拓扑关系模型主要包括 RCC 形式化模型和 n-交集模型。本文采用 9-交集模型^[17]区分 8 种常见的面/面拓扑关系,在 9-交集模型中,用对象的内部(A^0)、边界(∂A)和外部(A^-)子集是否相交来刻画两个对象间的拓扑关系:

$$T(A, B) = \begin{bmatrix} A^0 \cap B^0 & A^0 \cap \partial B & A^0 \cap B^- \\ \partial A \cap B^0 & \partial A \cap \partial B & \partial A \cap B^- \\ A^- \cap B^0 & A^- \cap \partial B & A^- \cap B^- \end{bmatrix}, \quad (1)$$

从理论上讲,由于集合有相交(用 1 表示)和不相交(用 0 表示)两种情况,因而 9-交集模型可区分 512 种关系,但其能区分的有意义的面/面拓扑关系是 8 种,如图 1 所示。

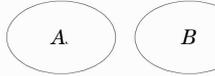
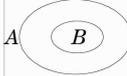
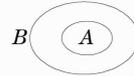
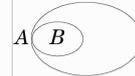
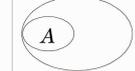
 $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ <p>disjoint</p>	 $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ <p>contain</p>	 $\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$ <p>inside</p>	 $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$ <p>cover</p>
 $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ <p>meet</p>	 $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ <p>equal</p>	 $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ <p>overlap</p>	 $\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$ <p>coveredBy</p>

图 1 八种拓扑关系及其九交模型

Fig. 1 Eight topological relations and their corresponding 9-intersection matrices

空间方向关系用于描述对象间的空间顺序。方向关系描述的方法有锥形方法、投影方法、最小边界矩形(MBR)方法、方向关系矩阵等。这里采用方向关系矩阵^[18]的方法,采用 MBR 来划分参照对象 A 的方向区域,将整个空间划分为 9 个子区域,包括东(E)、南(S)、西(W)、北(N)、东南(SE)、西南(SW)、西北(NW)、东北(NE)和一个中心区域(O),如图 2 所示,目标对象 B 与参照对象 A 之间的方向关系为 $\{NE_A, E_A\}$ 。

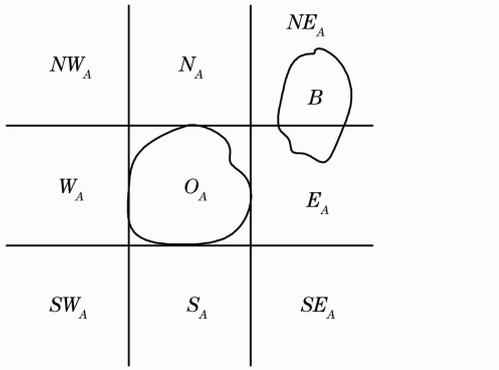


图 2 方向关系示意图

Fig. 2 Illustration of the directional relation

文献[18]用一个 3×3 的矩阵来定义方向关系矩阵,矩阵中的每个元素表示目标区域与方向区域相交情况。详细方向关系矩阵的每个元素的值域是 $[0, 1]$,以目标对象落在参考对象的各个方向区域的面积与目标对象总面积的百分比来表示方向关系

$$D(A, B) = \begin{bmatrix} \frac{S(NW_A \cap B)}{S(B)} & \frac{S(N_A \cap B)}{S(B)} & \frac{S(NE_A \cap B)}{S(B)} \\ \frac{S(W_A \cap B)}{S(B)} & \frac{S(O_A \cap B)}{S(B)} & \frac{S(E_A \cap B)}{S(B)} \\ \frac{S(SW_A \cap B)}{S(B)} & \frac{S(S_A \cap B)}{S(B)} & \frac{S(SE_A \cap B)}{S(B)} \end{bmatrix}, \quad (2)$$

式中 $D(A, B)$ 是详细方向关系矩阵,能够描述一定的数量信息,从数量上对定性方向关系作了进一步的区分,描述能力较强,但是该描述模型容易受位置不确定性影响,而且其描述仍然有别于人们的日常思维,如图 2 所示,人们一般描述其为“B 在 A 的东北方向”,即采用九种方向区域中的一种来描述其方向关系。因此,仅采用上式计算结果中的最大项对应的方向关系作为最后结果。

2.2 空间语义提取

根据实际经验知识可知人目标之间的空间语义包括前面(s_1)、后面(s_2)、背对(s_3)、面对(s_4)、左右(s_5)等 5 种:“前面”指目标对象在参考对象的前面(即目标对象背对着参考对象);“面对”指目标对象与参考对象面对面;“后面”是目标对象在参考对象的后面(即参考对象背对着目标对象);“背对”指目标对象与参考对象背对背。为了提取空间语义,本文定义一种空间特性 $F(i, j)$,由参考对象 i 与目标对象 j 之间的拓扑关系 $T(i, j)$ 、方向关系 $D(i, j)$ 、以及它们的运动方向 o_i 和 o_j 组成:

$$F(i, j) = [T(i, j), D(i, j), o_i, o_j]. \quad (3)$$

表 1 拓扑关系与方向关系的有效组合

Table 1 Available combination of topological and directional relation

	O	N	S	W	E	NE	SW	SE	NW
Disjoint	×	✓	✓	✓	✓	✓	✓	✓	✓
Touch	×	✓	✓	✓	✓	✓	✓	✓	✓
Overlap	✓	✓	✓	✓	✓	✓	✓	✓	✓

对于自然场景图像中的人目标,之前提到的 8 种拓扑关系中只有 Disjoint、Touch 和 Overlap 是有意义的。对于不同的拓扑关系,目标之间可能存在的方向关系也有所不同,如表 1 所示,其中有意义的空间关系组合共 25 种。量化目标运动方向如图 3 所示,则上式定义的四维的空间特性,共 $25 \times 8 \times 8 = 1600$ 种可能,根据实际经验,1600 种空间特性对应以上 5 种空间语义。表 2 列举几组有关空间特性分类空间语义的实例。

表 2 空间语义提取部分实例

Table 2 Examples about extraction of the spatial semantic meaning

(i, j)	$(T(i, j), D(i, j))$	O_i	O_j	Semantic meaning
1	(Dis., N)	D3	D3	S_1
2	(Dis., N)	D7	D7	S_2
3	(Dis., N)	D7	D3	S_3
4	(Dis., N)	D3	D7	S_4
5	(Dis., N)	D1	D1	S_5
6	(Dis., NW)	D3	D3	S_1
7	(Dis., NW)	D1	D1	S_2
8	(Dis., NW)	D7	D3	S_3
9	(Dis., NW)	D3	D7	S_4
10	(Dis., NW)	D2	D2	S_5

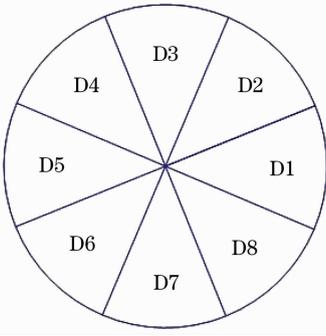


图 3 目标运动方向的量化

Fig. 3 Quantification of the motion direction of target

3 基于随机文法识别交互行为

根据运动目标跟踪结果,提取出参考对象与目标对象之间的空间语义,基于空间语义变化规律,采用随机文法器识别如下常见的两人交互行为类别:

x_1 : 两人同方向进入场景,其中一人一直跟随另一人;

x_2 : 两人同方向进入场景,最初一人跟随另一人,之后加速超过前人;

x_3 : 两人同方向进入场景,最初一人跟随另一人,随后加速追上前人,最后两人沿同一方向一起走;

x_4 : 两人同方向进入场景,最初一人跟随另一人,随后加速追上前人,随后一人返回,而另一人沿原来方向前进;

x_5 : 两人沿相反方向进入场景,之后相互接近、相遇,最后各自保持原来运动方向,两人相互远离;

x_6 : 两人沿相反方向进入场景,之后相互接近、相遇,最后两人沿同一方向前进;

x_7 : 两人同方向进入场景,一直并排走下去;

x_8 : 两人同方向进入场景,最初并排走,随后两人距离越来越大;

x_9 : 两人同方向进入场景,最初并排走,随后一个人返回。

随机文法定义成五元组 $G_s = (V_N, V_T, R, P, S)$, 式中, V_N 和 V_T 分别为非终止符集和终止符集, $S \in V_N$ 为起始符, R 为一组产生式规则, P 是概率集, R 的形式为

$$r: \alpha_i \xrightarrow{p_{ij}} \beta_{ij}, j = 1, 2, \dots, n; i = 1, 2, \dots, k, (4)$$

式中 $\alpha_i \in V^+$, 其中至少应有一个非终止符, $\beta_{ij} \in V^*$, 而 p_{ij} 是与这条随机生成式相联系的概率, i 指生成语义事件过程中的第 i 级生成, 而 j 是指第 i 级生成中第 j 级生成式, 显然有: $0 < p_{ij} \leq 1$ 和 $\sum_{j=1}^{n_i} p_{ij} = 1$ 。

定义随机文法五元组如下:

1) 非终止符 $V_N = \{S, A_1, A_2, B_1, B_2, C_1\}$, A_1, A_2, B_1, B_2, C_1 是变量, 并没有确切的含义, 其作用体现在表 3 中;

2) 起始符 S 为参考对象和目标对象的跟踪结果——运动轨迹;

3) 终止符 $V_T = \{s_1, s_2, s_3, s_4, s_5\}$, 包括前面 s_1 、后面 s_2 、背对 s_3 、面对 s_4 、左右 s_5 等空间语义;

4) 根据先验知识建立文法规则 R , 分析以上九种典型交互行为, 同时假设其先验概率相等, 如表 3 所示;

表 3 文法规则 R 及交互行为语法

Table 3 Rule production R and grammar of the interactions

ID	Production rule R	Probability	Grammar	Probability of the interaction
1	$S \rightarrow s_2$	p_1	$x_1 = s_2$	$p(x_1) = p_1 = 1/9$
2	$S \rightarrow s_5$	p_2	$x_7 = s_5$	$p(x_7) = p_2 = 1/9$
3	$S \rightarrow s_2 A_1$	p_3	—	—
4	$A_1 \rightarrow s_5$	p_6	$x_3 = s_2 s_5$	$p(x_3) = p_3 p_6 = 1/9$
5	$A_1 \rightarrow s_5 A_2$	p_7	—	—
6	$A_2 \rightarrow s_1$	p_8	$x_2 = s_2 s_5 s_1$	$p(x_2) = p_3 p_7 p_8 = 1/9$
7	$A_2 \rightarrow s_4$	p_9	$x_4 = s_2 s_5 s_4$	$p(x_4) = p_3 p_7 p_9 = 1/9$
8	$S \rightarrow s_3 B_1$	p_4	—	—
9	$B_1 \rightarrow s_5$	p_{10}	$x_6 = s_4 s_5$	$p(x_6) = p_4 p_{10} = 1/9$
10	$B_1 \rightarrow s_5 B_2$	p_{11}	—	—
11	$B_2 \rightarrow s_4$	p_{12}	$x_5 = s_3 s_5 s_4$	$p(x_5) = p_4 p_{11} p_{12} = 1/9$
12	$S \rightarrow s_5 C_1$	p_5	—	—
13	$C_1 \rightarrow s_1$	p_{13}	$x_8 = \{s_5 s_1,$	$p(x_8) = p_5 p_{13} + p_5 p_{14} = 1/9$
14	$C_1 \rightarrow s_2$	p_{14}	$s_5 s_2\}$	
15	$C_1 \rightarrow s_4$	p_{15}	$x_9 = s_5 s_4$	$p(x_9) = p_5 p_{15} = 1/9$

5) 概率 $P(p_1 \sim p_{15})$ 是与其相应的随机生成式联系的概率) 的约束条件包括: $p_1 + p_2 + p_3 + p_4 + p_5 = 1, p_6 + p_7 = 1, p_8 + p_9 = 1, p_{10} + p_{11} = 1, p_{12} = 1, p_{13} + p_{14} + p_{15} = 1$ 。

求解以上方程及表 3 中的方程, 可得: $p_1 = 1/9, p_2 = 1/9, p_3 = 1/3, p_4 = 2/9, p_5 = 2/9, p_6 = 1/3, p_7 = 2/3, p_8 = 1/2, p_9 = 1/2, p_{10} = 1/2, p_{11} = 1/2, p_{12} = 1, p_{13} = 1/4, p_{14} = 1/4, p_{15} = 1/2$ 。

4 实验与分析

为了验证方法的有效性, 采用 CASIA 数据库^[19]中的五种交互行为视频以及拍摄的其他四种交互行为视频各 12 段。运动目标的检测与跟

踪^[20~24]是交互行为识别的前提, 采用文献[20]中的方法检测并跟踪运动目标, 并实时提取目标之间的空间语义信息, 根据空间语义的变化情况, 采用前面建立的随机文法器识别目标交互行为。

4.1 交互行为识别

图 4 为采用本文方法识别交互行为 x_2 和 x_5 的分析图, 通过目标检测和跟踪, 获取参考对象和目标对象的运动方向 (Ref., Pri.), 建模其空间关系 (拓扑关系 Top. 和方向关系 Dir.), 根据 2.2 节的内容实时提取其空间语义 (Sem.), 最后基于空间语义的变化采用随机文法器识别其交互行为。如图 4(a) 所示, 可得其行为语法为 $x = s_2 s_5 s_1$, 同时可判断其行为类别为 x_2 ; 如图 4(b) 所示, 可得其行为语法为 $x = s_3 s_5 s_4$, 同时可判断其行为类别为 x_5 。

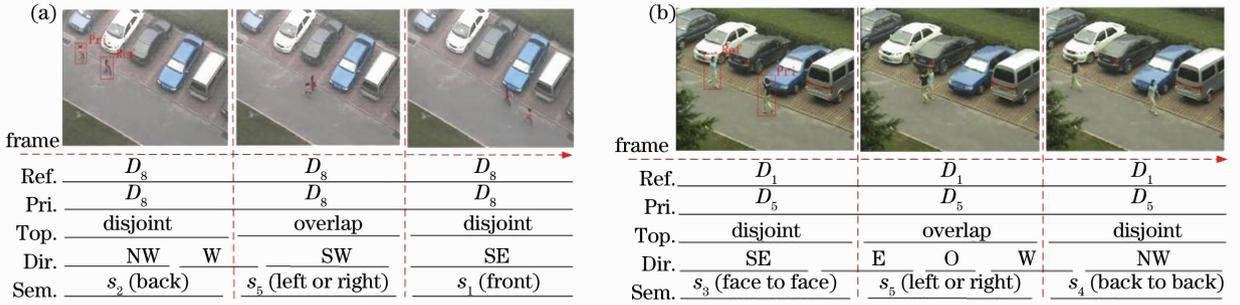


图 4 交互行为识别实例。(a) $x_2 = s_2 s_5 s_1$; (b) $x_5 = s_3 s_5 s_4$

Fig. 4 Examples of the interactions. (a) $x_2 = s_2 s_5 s_1$; (b) $x_5 = s_3 s_5 s_4$

4.2 对比实验

表 4 为采用提出的方法、HMM^[7]、CHMM^[7]、VLMM^[10]以及 MTP^[14]在测试数据集上的实验性能对比, 从中可以看出, 提出的方法性能远远优于文献中的算法。运动目标检测与跟踪是进行行为识别的前提, 然而相对于传统方法通过建模运动轨迹识别交互行为类别, 本文算法的稳健性较强。算法采用的空间语义信息的提取方法简单, 目标检测与跟

踪结果的不稳定 (如漏检、跟踪丢失等) 对空间语义的变化规律的影响较小, 如假设两目标之间真实的空间语义信息序列是 $s_2 s_2 s_2 s_2 s_5 s_5 s_5 s_5 s_1 s_1 s_1$, 由于跟踪丢失造成实际提取的语义列表是 $s_2 s_2 s_2 s_5 s_5 s_1 s_1$, 但是其空间语义变化规律都是 $s_2 s_5 s_1$, 不影响该交互行为的文法语法, 可以正确识别其交互行为类别。空间语义序列可以反映时间特性, 同时空间语义信息的变化规律可以有效地表征其交互行为类别。

表 4 新方法与传统方法的性能对比

Table 4 Performance comparisons of our algorithm with other methods

Interaction	HMM / %	CHMM / %	VLMM / %	MTP / %	Ours / %
x_1	83.4	75	83.3	75	100
x_2	50	58.3	66.7	75	100
x_3	75	75	83.3	83.3	100
x_4	58.3	66.7	91.7	100	100
x_5	66.7	83.3	100	100	100
x_6	75	100	100	100	100
x_7	83.3	75	75	83.3	91.7
x_8	50	83.4	83.4	91.7	100
x_9	58.4	75	75	83.3	100
Average	66.7	76.9	84.3	88	99.1

图 5 为分别采用本文方法、HMM^[7]、CHMM^[7]、VLMM^[10]以及 MTP^[14]识别九种交互行为类别的混淆矩阵,从中可以看出,新的方法比传

统方法的性能更稳定,主要原因是空间语义的变化规律较原子事件或运动轨迹等传统特征更有效、更稳定。

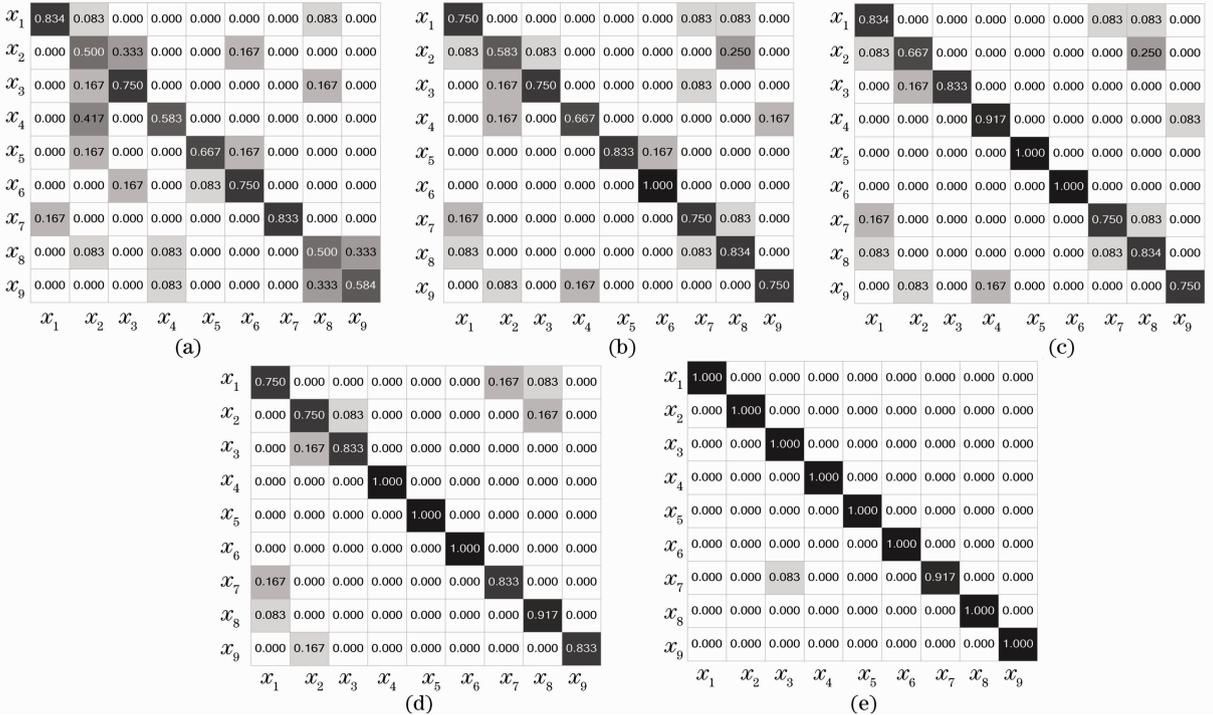


图 5 交互行为识别混淆矩阵。(a) HMM; (b) CHMM; (c) VLMM; (d) MTP; (e) 新方法

Fig. 5 Confusion matrices with different approaches in recognizing multi-agent interactions. (a) HMM; (b) CHMM; (c) VLMM; (d) MTP; (e) proposed approach

5 结 论

提出了一种有关视频人目标交互行为识别方法,不同于传统识别方法,该方法融合时间和空间语义信息,基于空间语义变化规律建立随机文法器识别交互行为,不需要训练样本,空间语义信息序列反映了其时间特性,空间语义信息的变化规律可以有效地表征其交互行为类别。实验结果验证了采用的空间语义特性优于传统的简单原子事件。

提出的方法在九种典型的两人交互行为识别实验中取得了较好的效果。今后可以改进本方法,用于识别较复杂的交互行为,如握手、拥抱和打架等,首先提取目标部件之间的空间语义信息,如四肢与人体躯干之间的上下左右关系等,通过挖掘交互行为类别与空间语义信息的变化规律之间的关系,进而识别其交互类别。同时可以采用提出的空间语义信息用于建模并识别复杂场景中的多人交互行为。

参 考 文 献

1 G. Lavee, E. Rivlin, M. Rudzsky. Understanding video events;

a survey of methods for automatic interpretation of semantic occurrences in video [J]. *IEEE Trans. Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 2009, **39**(5): 489~504

2 M. S. Ryoo, J. K. Aggarwal. Human activity analysis; a review [J]. *ACM Computer Surveys*, 2011, **43**(3): 16

3 Li Hongsong, Li Da. Some advances in human motion analysis [J]. *Pattern Recognition and Artificial Intelligence*, 2009, **22**(1): 70~78

黎洪松, 李 达. 人体运动分析研究的若干新进展 [J]. *模式识别与人工智能*, 2009, **22**(1): 70~78

4 J. W. Davis, A. F. Bobick. The representation and recognition of human movement using temporal templates [C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'97), Washington, DC, USA, 1997: 928

5 M. S. Ryoo, J. K. Aggarwal. Spatio-temporal relationship Match; video structure comparison for recognition of complex human activities [C]. In Proceedings of the IEEE International Conference on Computer Vision (ICCV'09), Kyoto, Japan, Oct 2009

6 S. Hongeng, R. Nevatia, F. Bremond. Video-based event recognition; activity representation and probabilistic recognition methods [J]. *Computer Vision and Image Understanding (CVIU)*, 2004, **96**(2): 129~162

7 N. M. Oliver, B. Rosario, A. P. Pentland. A Bayesian computer vision system for modeling human interactions [J]. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2000,

- 22(8); 831~843
- 8 S. Park, J. K. Aggarwal. A hierarchical Bayesian network for event recognition of human actions and interactions [J]. *Multimedia System*, 2004, **10**(2): 164~179
- 9 P. Natarajan, R. Nevatia. Coupled hidden semi Markov models for activity recognition [C]. In IEEE Workshop on Motion and Video Computing (WMVC'07), Austin, TX, USA, 2007
- 10 A. Galata, A. G. Cohn, D. Magee *et al.*. Modeling interaction using learned qualitative spatio-temporal relations and variable length Markov models [C]. Proceeding of the 15th European Conference on Artificial Intelligence (ECAI'02), 2002: 741~745
- 11 Han Lei, Li Junfeng, Jia Yunde. Human interaction recognition using spatial-temporal words [J]. *Chinese J. Computers*, 2010, **33**(4): 776~784
韩磊, 李君峰, 贾云得. 基于时空单词的两人交互行为识别方法 [J]. *计算机学报*, 2010, **33**(4): 776~784
- 12 M. S. Ryoo, J. K. Aggarwal. Semantic representation and recognition of continued and recursive human activities [J]. *Internat. J. Computer Vision*, 2009, **82**(1): 1~24
- 13 M. Sridhar, A. G. Cohn, D. C. Hogg. Unsupervised learning of event classes from video [C]. Proc. AAAI, 2010: 1631~1638
- 14 Z. Zhang, T. N. Tan, K. Q. Huang. An extended grammar system for learning and recognizing complex visual events [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **33**(2): 240~255
- 15 S. W. Joo, R. Chellappa. Attribute grammar-based event recognition and anomaly detection [C]. In Proceedings of 2006 Conference on Computer Vision and Pattern Recognition Workshop, 2006: 107
- 16 M. S. Ryoo, J. K. Aggarwal. Stochastic representation and recognition of high-level group activities [J]. *Internat. J. Computer Vision*, 2011, **93**(2): 183~200
- 17 M. J. Egenhofer, R. D. Franzosa. Point-set topological spatial relations [J]. *Internat. J. Geograph. Inform. Systems*, 1991, **5**(2): 161~174
- 18 R. Goyal. Similarity Assessment for Cardinal Directions Between Extended Spatial Objects [D]. Maine: University of Maine, 2000
- 19 The Center for Biometrics and Security Presearch. CASIA action Database [DB/OL]. <http://www.cbsr.ia.ac.cn/english/Databases.asp>
- 20 Jin Biao, Hu Wenlong, Wang Hongqi. Occlusion handling method for multiple moving objects tracking based on multilevel tracker queues [J]. *Acta Optica Sinica*, 2011, **31**(8): 0815001
金标, 胡文龙, 王宏琦. 基于多级跟踪队列的运动目标跟踪遮挡处理 [J]. *光学学报*, 2011, **31**(8): 0815001
- 21 Liang Min, Liu Guixi. Multiobject tracking algorithm based on adaptive mixed filtering [J]. *Acta Optica Sinica*, 2010, **30**(9): 2554~2561
梁敏, 刘贵喜. 基于自适应混合滤波的多目标跟踪算法 [J]. *光学学报*, 2010, **30**(9): 2554~2561
- 22 Zhang Xuguang, Zhang Yun, Wang Yanning *et al.*. Covariance tracking based on forgetting factor and Kalman filter [J]. *Acta Optica Sinica*, 2010, **30**(8): 2317~2323
张旭光, 张云, 王艳宁等. 基于遗忘因子与卡尔曼滤波的协方差跟踪 [J]. *光学学报*, 2010, **30**(8): 2317~2323
- 23 Gao Lin, Tang Peng, Sheng Peng *et al.*. Visual object tracking based on conditional random field under complex scene [J]. *Acta Optica Sinica*, 2010, **30**(6): 1721~1728
高琳, 唐鹏, 盛鹏等. 复杂场景下基于条件随机场的视觉目标跟踪 [J]. *光学学报*, 2010, **30**(6): 1721~1728
- 24 Shen Huan, Li Shunming, Bo Fangchao *et al.*. On road vehicles real-time detection and tracking using vision based approach [J]. *Acta Optica Sinica*, 2010, **30**(4): 1076~1083
沈岷, 李舜韶, 柏方超等. 路面车辆实时检测与跟踪的视觉方法 [J]. *光学学报*, 2010, **30**(4): 1076~1083

栏目编辑：李文喆