

# 复杂背景下跟踪多个非刚体 运动目标的快速算法

王方石<sup>1)</sup> 须德<sup>2)</sup>

1) (北京交通大学计算机与信息技术学院计算机科学与技术系, 北京 100044)

2) (北京交通大学计算机与信息技术学院计算机研究所, 北京 100044)

**摘要** 在复杂背景下对多个非刚体运动目标进行跟踪是视频对象分割中的一个难点, 为此提出了一种跟踪多个非刚体运动目标轮廓并提取精确视频对象平面(VOP)的半自动快速算法。手工获得初始模板后, 利用边界投射模板、Canny算子和 Hausdorff 距离匹配算法, 可快速、有效地自动跟踪多个非刚体的运动目标。此外还提出了一种新方法, 较好地解决了被遮挡背景的显露问题, 并能得到任意形状的闭合轮廓线, 继而提取精确的 VOP。该算法取得了较好的实验结果。

**关键词** 多个非刚体运动目标 Canny 算子 Hausdorff 距离匹配 边界投射模板 闭合轮廓线 VOP

中图分类号: TP391 TN941.1 文献标识码: A 文章编号: 1006-8961(2005)06-0746-08

## A Fast Algorithm of Tracking Multi-nonrigid Moving Objects in the Clutter Background

WANG Fang-shi<sup>1)</sup>, XU De<sup>2)</sup>

<sup>1)</sup> (Department of Computer Science & Technology, School of Computer & Information Technology, Beijing Jiaotong University, Beijing 100044)

<sup>2)</sup> (Institute of Computer, School of Computer & Information Technology, Beijing Jiaotong University, Beijing 100044)

**Abstract** It is a key problem to track Multi-nonrigid moving objects in the clutter background and extract the video object planes (VOP's) in video object segmentation. This paper presents a fast semi-automatic algorithm of tracking multi-nonrigid moving object and extracting the accurate video object planes. After obtaining the initial mask manually, it is able to track multi-nonrigid moving objects fast, efficiently and automatically. First, it obtains the Edge Projection Mask from the initial mask using Block Matching algorithm. Then Hausdorff Distance Matching is done between Change Detection Mask and Gradient picture of the next frame. Third, it combines the results of the first two steps and gets more integrated contours of the objects. Additionally, a new method is also presented, which can not only solve the uncovered background better, but also close the contours of the objects with convex or concave shape and extract the accurate VOP's from the sequence. Experimental results demonstrate the performance of our proposed algorithm.

**Keywords** multi-nonrigid moving objects, Canny operator, Hausdorff distance match, edge projection mask, closed contour, VOP

## 1 引言

MPEG-4 规定视频由有意义的视频对象组成, 在编码前需将视频序列分割为语义意义上的视频对

象(VO)或视频对象平面(VOP)。因此人们对视频对象分割算法进行了深入的研究<sup>[1]</sup>。目前视频对象分割方法大致分为自动分割和半自动分割两类。

自动分割算法存在着许多缺点, 如只能分割一些很简单的语义对象; 当目标和背景具有相似颜色

收稿日期: 2004-10-08; 改回日期: 2004-12-16

第一作者简介: 王方石(1969~), 女, 副教授。1990年于吉林大学计算机系获学士学位, 1993年于吉林大学数学研究所获硕士学位, 目前是北京交通大学计算机应用专业的在职博士研究生。主要研究方向为基于内容的视频检索、模式识别。E-mail: wfs@computer.njtu.edu.cn

和纹理时,无法获得精确的目标边界。文献[2]提出的自动分割目标,先用边界检测算子得到边界点集合,记作 *edge*,再采用区域增长算法得到区域边界点集合,记作 *boundary*,然后根据 *edge* 和 *boundary* 中点的位置关系确定应保留的目标边界,其优点是所得到的目标轮廓是闭合的,然而,当目标和背景具有相似颜色和纹理时,无论是边界检测算子还是区域增长算法,都无法找到应有的目标边界点,导致分割不准确。半自动分割方法通过人机交互弥补了计算机在语义理解上的缺陷,解决了计算机无法界定模糊边缘的难题。

所有跟踪算法都经过3个阶段:获得初始目标模板、目标跟踪和提取目标 VOP。在获得初始目标模板阶段,文献[3]采用差分图像和初始帧梯度图像的“与”操作得到初始模板,然后采用活动轮廓模型闭合边界轮廓(Snake),其前提是显露的背景中没有明显的边缘,而且 Snake 并不能完全解决在图像中寻找轮廓的问题,有时仍需依靠其他机制,比如与用户交互,由用户指定轮廓估计的形状和开始位置,或应用更高级的图像理解处理方法<sup>[4]</sup>,处理复杂且计算量大,而且当图像中前景、背景颜色纹理相近时,处理效果不理想。文献[5]、[6]采用独立运动部件和初始帧梯度图像的距离变换,获得初始模板,它未闭合,并且不十分精确,这种误差会在目标跟踪阶段被放大,从而得到后续帧中不精确的 VOP。

在目标跟踪阶段,因目标移动而显露的背景纹理常常会纳入目标模板。为此,文献[3]、[5]、[6]均采用计数法,设置一个累积差分(计数)缓冲区,考察前  $N$  帧(文献[3]中  $N=8$ )每相邻两帧差分模板,统计该像素在差分模板中为1的次数,若次数为0,则视为背景点,否则视为前景点,此方法必须连续读入  $N$  帧后才能判断某像素是否为背景,而且实验结果显示:该方法仍无法完全剔除运动目标附近的背景像素。

在提取目标 VOP 阶段,首先要闭合轮廓线。文献[3]仍采用活动轮廓模型闭合边界。文献[5]则提出了二次行扫描和一次列扫描的简单填充算法,这种方法仅适用于凸轮廓的运动目标,对于凹轮廓的运动目标,会把不属于对象的背景纳入该目标,如图1(a)所示,人两腿之间的地板被当成人的一部分了。对其加以改进,采用 Dijkstra 最短距离算法修正错误边界<sup>[6]</sup>,但结果仍不十分精确,有毛刺及背景像素,如图1(b)所示。

为了消除毛刺,许多算法采用数学形态学的开



(a) 文献[5]的结果

(b) 文献[6]的结果

图1 文献结果

Fig.1 VOP extracted in paper [5] and [6]

算子消除毛刺,其缺点是要根据具体应用设定结构元素的大小和形状。采用含有256种情况的查找表算法,其缺点是需事先构造查找表<sup>[6]</sup>。

针对上述算法中存在的问题,提出一种准确快速的跟踪非刚体运动目标的算法。首先采用简单的人机交互操作,解决精确分割颜色纹理相近的前景和背景的问题,然后自动跟踪后续帧中的运动目标。同时还提出一种简单的新算法,可剔除背景像素,快速地闭合目标轮廓线,最终获得多个目标的精确 VOP。

## 2 目标初始模板的形成

采用简单的半自动方法提取初始帧的目标模板。如图2所示,首先采用 Canny 算子得到初始帧的梯度图像(图2(c)),然后通过帧间差得到运动目标的大致位置,形成二值差分模板 CDM (change detection mask, 也称变化检测模板,如图2(d)所示),对两者进行 Hausdorff 距离匹配,得到未闭合的二值模板(图2(e)),当前景、背景相近时,边缘是残缺的,将它叠加在初始帧的原图像上,以方便用户手工弥补未闭合的部分。

### 2.1 边缘检测

一般用图像的空间梯度来表示边缘信息。求取边缘的梯度算子很多,例如 Robert 算子、Sobel 算子、Prewitt 算子和形态梯度算子等。由于噪声的存在,通常要对图像进行滤波处理,但滤波也会使边缘变得模糊,因此,上述方法得到的并不是精确的边缘,而是具有一定宽度。为了得到单像素宽度的精确的对象边缘,对初始帧采用 Canny 算法求得梯度图像  $G$ ,如图2(c)所示。

### 2.2 运动检测

采用基于 CDM 的方法,进行运动检测,它适用于运动较快的非刚体目标<sup>[7]</sup>。

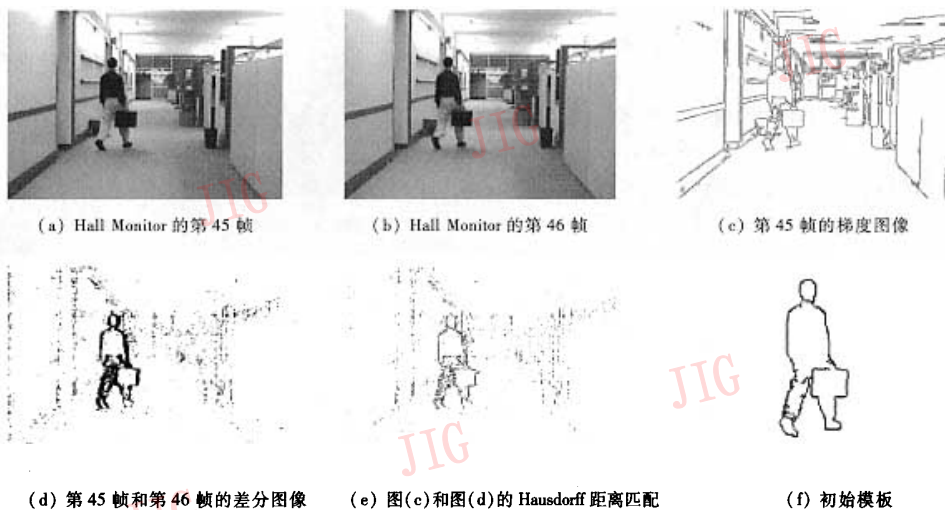


图2 获得 Hall Monitor 中初始模板的过程

Fig.2 The process of obtaining the initial mask

(1) 若存在全局运动,则先对背景做运动补偿。

(2) 对连续两帧(第  $t$  帧和第  $t+1$  帧)图像做帧差,得到二值变化检测模板  $CDM$ (也称差分图像),其中既包含真正的运动目标,也包含噪声及由于目标运动而显露出来的背景,如图 2(d) 所示。

### 2.3 模板的初始化

对梯度图像和差分图像进行 Hausdorff 距离匹配,得到初始边缘模板  $E_i$ ,

$$E_i = \left\{ e \in G_i \mid \min_{x \in CDM} \|x - e\| \leq T_1 \right\} \quad (1)$$

其中,  $T_1$  取值为 0,即初始边缘模板由同时属于  $CDM$  和  $G_i$  中的所有边界点构成。图 2(e) 是图 2(c) 和图 2(d) 进行 Hausdorff 距离匹配的结果。实际上,变化检测模板中不仅包含运动的物体,还包含有显露出来的背景部分,如图 2(d) 所示,由于人是从左向右移动,人左侧有宽边缘,但在图 2(e) 中,只保留了人体的边缘轮廓,可见,空间梯度与时间梯度相结合,在一定程度上可自动消除显露背景,但要求显露的背景中没有明显的边缘,否则效果也不理想。

#### (1) 闭合轮廓

通过上述操作得到的目标轮廓往往不是闭合的。若自动闭合轮廓线,会将目标附近的背景纹理纳入目标边界中,从而造成模板不精确的效果,误差也会传递到目标跟踪阶段。

本系统采用手工闭合方式,将前述得到的初始边缘模板  $E_i$  叠加在第 1 帧原图像上,以使用户手工

勾画,同时将背景纹理轮廓与目标轮廓线断开。

#### (2) 提取初始帧的 VOP

至此,边缘模板  $E_i$  中除了闭合的轮廓线,还有因纹理复杂而保留下来的多余的边界点,为消除它们,可采用滤除噪声的方法进行处理,这里采用区域增长算法,统计  $E_i$  中每个连通分量所含有的像素点数,若点数超过 50,认为该区域不是噪声,保留;若不超过 50,就将该连通分量整体滤除。这样便可消除与目标轮廓线分离的背景边缘。然后对背景进行区域增长,可得到目标轮廓点的集合,记为  $B$ ,如图 2(f) 所示。

## 3 模板更新

视频对象在视频序列中会产生运动,有些甚至发生形变。模板更新就是根据前一帧运动目标的模板,构造出当前帧中运动目标的修正模板。仅从当前帧得到的边缘模板往往残缺不全,为了得到完整的目标轮廓,必须借助前一帧目标模板来修正当前帧的目标模板,从而得到精确的新目标模板。

### 3.1 运动估计

若背景非静止,先进行全局运动估计,然后采用文献[8]中的方法得到局部运动矢量。对前一帧目标模板进行运动补偿,得到下一帧的预测模板。若背景静止,则对连续两帧采用块匹配算法,前一帧边界点集  $B$  在下一帧中的最佳匹配位置构成了边界投射模板  $M_{t,t+1}$ (如图 3(a) 所示)。

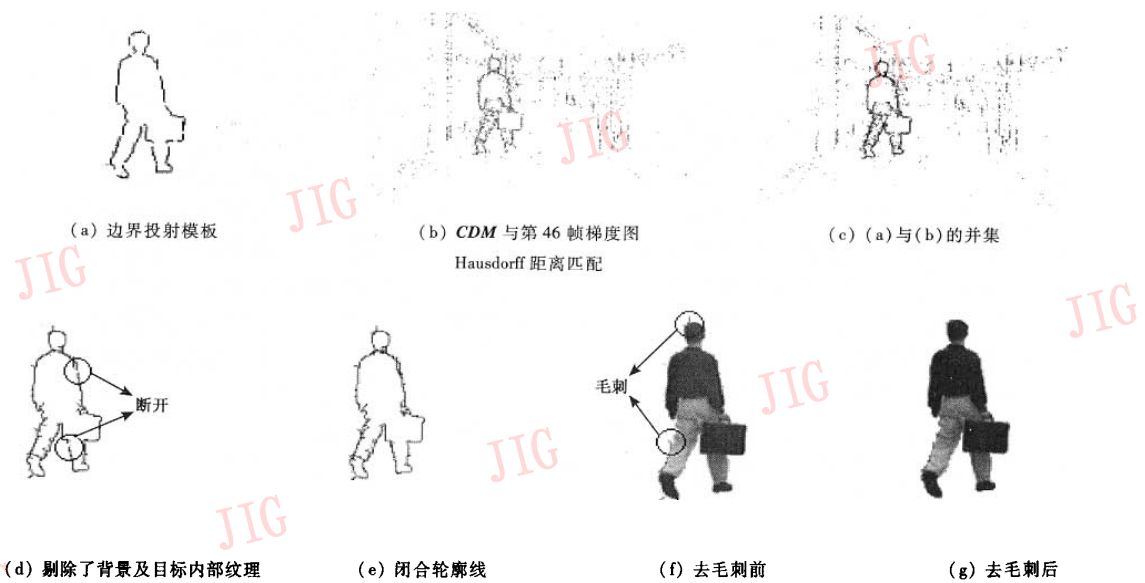


图 3 由 Hall Monitor 中初始模板获得下一帧 VOP 的过程

Fig. 3 The process of obtaining VOP in the next frame from the initial mask in Hall Monitor

$$M_{t+1} = \{p \oplus v_p | p \in B \text{ and } v_p \in V\} \quad (2)$$

其中,  $v_p$  是像素点  $p$  的运动矢量,  $p \oplus v_p$  表示  $p$  经过  $v_p$  平移到当前帧中最佳匹配位置。在求  $v_p$  时, 为了提高运算速度, 只对含有边界点的少数块进行前向块匹配, 为保证匹配精度, 采用全搜索方式, 既保证了运动估计的准确性, 也保证了较高的运算速度。

### 3.2 当前帧中目标的粗略模板

首先对当前帧进行 Canny 算子操作, 得到当前帧的梯度图像  $G_{t+1}$ , 为去掉其中包含的复杂背景边界, 将  $G_{t+1}$  和差分图像  $CDM$  进行 Hausdorff 距离匹配, 得到  $E_{t+1}$ ; 然后求运动补偿得到的边界投射模板  $M_{t+1}$  与  $E_{t+1}$  的并集, 以弥补  $E_{t+1}$  中因前景和背景相近而造成的目标边界的缺口, 其过程如下:

(1) 将当前帧的梯度图像  $G_{t+1}$  和差分图像  $CDM$  进行 Hausdorff 距离匹配, 得到模板

$$E_{t+1} = \left\{ e \in G_{t+1} \mid \min_{x \in CDM} \|x - e\| \leq T_2 \right\} \quad (3)$$

其中,  $T_2 = 0$ ,  $E_{t+1}$  中既包含目标的真正边界, 也包含目标内部纹理和因目标运动而显露的背景纹理, 但目标轮廓未必完整, 如图 3(b) 所示, 仍保留了不必要的纹理像素, 且目标左右脚的轮廓残缺。

(2) 用边界投射模板  $M_{t+1}$  弥补当前帧残缺的轮廓, 求  $M_{t+1}$  与  $E_{t+1}$  的并集, 可得到较完整但未闭合的粗略模板, 记作  $C_{t+1}$ ,

$$C_{t+1} = E_{t+1} \cup M_{t+1} \quad (4)$$

如图 3(c) 所示, 目标左右脚的轮廓已变得很完整。

### 3.3 剔除背景像素或目标内部纹理

在实验中发现, 按照上面的更新方法做, 可以得到当前帧中较完整的边缘信息, 但也保留了目标附近的背景像素, 若不进行消除, 只能得到不精确的目标模板。

为此提出一种新算法, 以可消除目标内部纹理点, 并剔除其附近的背景像素。算法实现的过程如下:

(1) 消除远离目标边界的背景像素和目标内部的纹理点。从实验中可知, 运动补偿得到的边界一定在真正边界的附近, 因此附近没有  $M_{t+1}$  点的  $E_{t+1}$  点一定对应背景纹理或目标内部纹理, 应消除, 这类点可用下式描述

$$D_1 = \left\{ e \in E_{t+1} \mid \min_{x \in M_{t+1}} \|x - e\| \geq T_3 \right\} \quad (5)$$

其中,  $T_3 = 5$ 。

(2) 消除因块匹配算法不精确而偏离真正边界的运动补偿点。与  $E_{t+1}$  点相邻的  $M_{t+1}$  点一般是块匹配得到的不准确预测边界, 也应消除, 这类点可用下式描述

$$D_2 = \left\{ x \in M_{t+1} \mid \min_{e \in E_{t+1} \wedge e \neq x} \|x - e\| \leq T_4 \right\} \quad (6)$$

其中,  $T_4 = 1$ 。

附近没有  $E_{t+1}$  点的  $M_{t+1}$  点被认为是残缺边界的弥补, 应保留。

如此操作, 可达到简化图像模板的目的, 它是确

保系统正确闭合目标轮廓线的必要且关键的一步。此时当前帧中目标的模板,如图 3(d)所示。

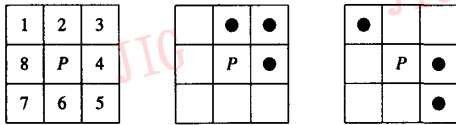
### 3.4 闭合运动目标的轮廓线

上述过程所得到的目标边界形状是准确的,但不一定能构成运动对象的闭合轮廓线,而且绝大多数情况下,是得不到的。

由于边界投射模板可以弥补因前景和背景相近而造成的目标边界缺失,此时的目标轮廓线上只有少数几个缺口,大小一般在 5 个像素宽度之内,可采用自动方式闭合轮廓线。这里提出了一种简单的新算法,可以快速有效地闭合轮廓线。系统自动闭合的轮廓线,如图 3(e)所示。具体步骤如下:

(1) 经过上述操作,运动补偿边界点附近还有些细碎的背景纹理,因此仍需采用 2.3 节第 2 步的方法滤除当前帧模板中背景噪声;

(2) 找出断点,断点定义为在其 3×3 邻域内与其相邻的前景点是连通的点。在目标模板中,用  $t(p) = 1$  表示像素  $p(x, y)$  是前景点,否则为背景点。将  $P(x, y)$  周围 3×3 邻域内的点按顺时针标号,如图 4(a)所示。



(a) 与 P 相邻点的标号 (b) P 是断点 (c) P 不是断点

图 4 判断断点

Fig. 4 Broken point or not

令  $t(9) = t(1)$ , 计算  $n(x, y)$  的值

$$n(x, y) = \frac{1}{2} \sum_{i=1}^9 |t(i) - t(i+1)| \quad (7)$$

$n(x, y) = 1$  说明  $p(x, y)$  是断点, 否则不是。如图 4(b)中的 P 点是断点, 而图 4(c)中的 P 点不是



(a) 《情定大饭店》第 35 帧



(b) 第 35 帧的初始模板



(c) 第 36 帧的 VOP



(d) 第 55 帧的 VOP

图 5 韩剧《情定大饭店》序列中第 35、36、55 帧的 VOP

Fig. 5 The VOPs of frame 35, 36, 55 in the Korean teleplay "Hotel"

(2) Hall Monitor 该序列每帧大小为 352 × 240, 由前一帧模板得到下一帧精确模板平均仅用 0.365s。

断点;

(3) 在以断点为中心的  $N \times N$  的邻域内找另一个断点, 并连接它们。在本系统中,  $N$  取值 11;

(4) 重复第 2、3 步, 直到在所有断点的  $N \times N$  邻域内找不到另一个断点为止;

(5) 消除毛刺。虽然经过 3.3 节和本节第 1 步的处理, 但只能消除与前景不连通的背景像素, 此时, 闭合轮廓线上还会附着着 1 个或 2 个像素宽的背景, 如图 3(f) 所示, 称之为毛刺。这里提出一种简单快速有效的消除毛刺算法: 重新计算  $n(x, y)$ , 由于目标轮廓线闭合后, 边界点的  $n$  值均大于 1, 若  $n(x, y) = 1$ , 说明它是断点, 即毛刺, 则消除它, 找到相邻的点重复此过程, 可快速有效地消除毛刺。图 3(f) 为消除毛刺前的目标模板, 图 3(g) 为消除毛刺后的目标模板。

### 3.5 提取 VOP

与 2.3 节第 2 步的方法相同。

## 4 实验结果分析

在 AMD Athlon 2500 +, 256MB 内存, Windows XP 环境下, 用 VC++ 编程实现了本算法。对不同特点的图像序列做了大量的测试, 限于篇幅, 仅以下列视频序列为例: 韩剧《情定大饭店》片断, Hall Monitor, 和电影《真爱至上》片断。程序在各阶段分别使用了 4 个阈值  $T_1, T_2, T_3, T_4$ , 其取值均为经验值, 而且针对所有的实验采用的是统一的值。

(1) 韩剧《情定大饭店》片断 该序列每帧大小为 448 × 336, 由前一帧模板得到下一帧精确模板平均用 0.937s。在该序列中, 两目标在对话时身体及头部均有运动, 连续提取的 VOP 如图 5 所示, 获得轮廓边界比较精确。

在该序列的第 125、126 帧中, 左侧目标由近到远, 右侧目标由远到近, 从图 6(i) 可见, 第 125 帧的边界投射模板弥补了第 126 帧的残缺轮廓线, 同时

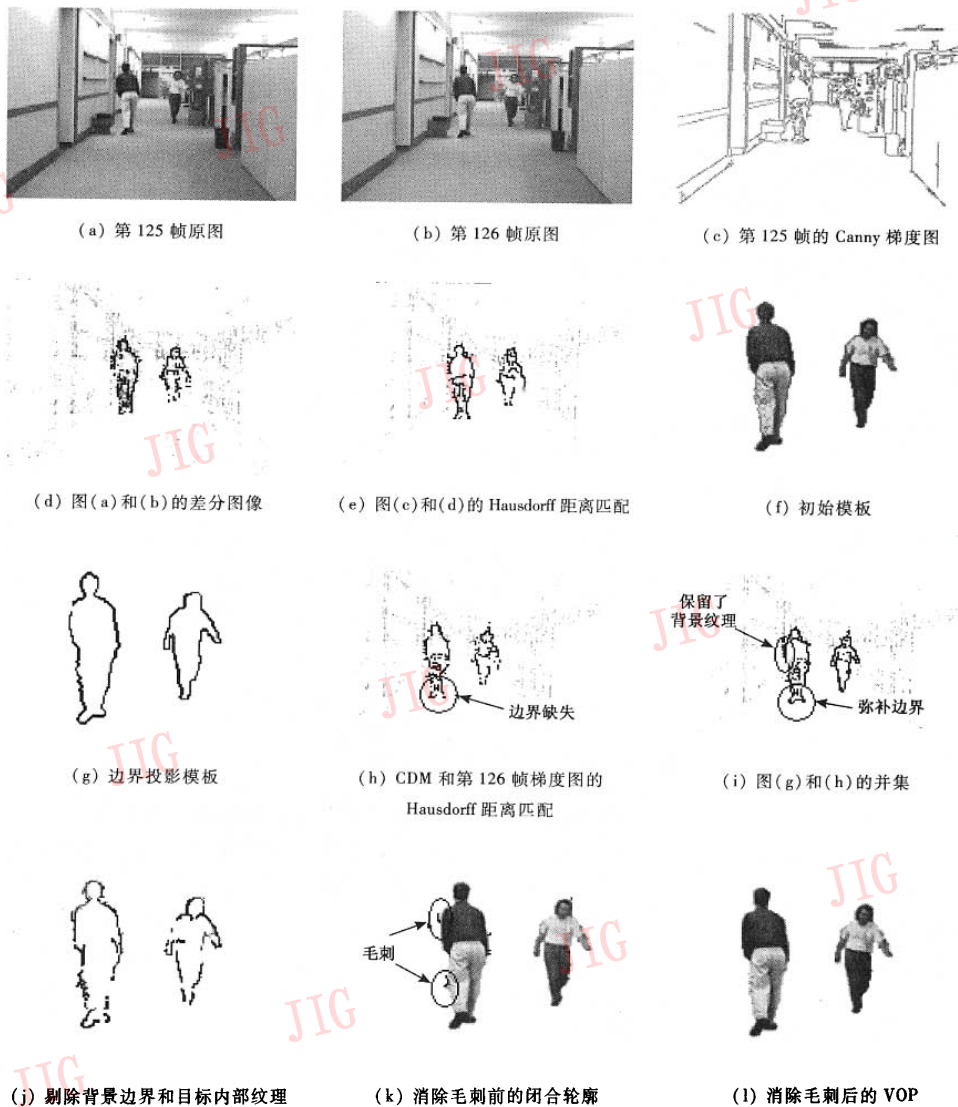


图6 Hall Monitor 序列提取 VOP 的过程

Fig. 6 The process of obtaining VOP in Hall Monitor

也发现,经过梯度图像和差分图像的 Hausdorff 距离匹配后,运动目标附近仍保留了显露的背景纹理。在闭合轮廓线前应消除它,否则系统也会处理这些非目标轮廓的边界点,不仅会降低算法效率,更糟糕的是还会将背景纹理连入目标轮廓,造成模板的不精确。采用本文算法成功地消除了显露背景纹理和目标内部纹理,如图 6(j) 所示。闭合轮廓线后,发现目标边界有毛刺,这是因为它们与真正的轮廓点相连通,采用 3.3 节的方法及滤除噪声的方法,均无法消除它们。对此,采用 3.4 节第 5 步,成功地消

除了这些毛刺,如图 6(l) 所示。

(3) 电影《真爱至上》片断 该序列每帧大小为  $640 \times 272$ ,由前一帧模板得到下一帧精确模板平均用 1.4s。

由于图中女士的衣服与背景墙的颜色相近,无论是帧差图像中,还是梯度图像中,女士左臂的轮廓线都缺失了大部分,如图 7(c) 所示。在提取初始帧的目标模板时,只能采用手工闭合轮廓线的方法;当已知前一帧模板提取下一帧目标轮廓时,就无需手工闭合轮廓线了,因为有前一帧边界投射模板可弥

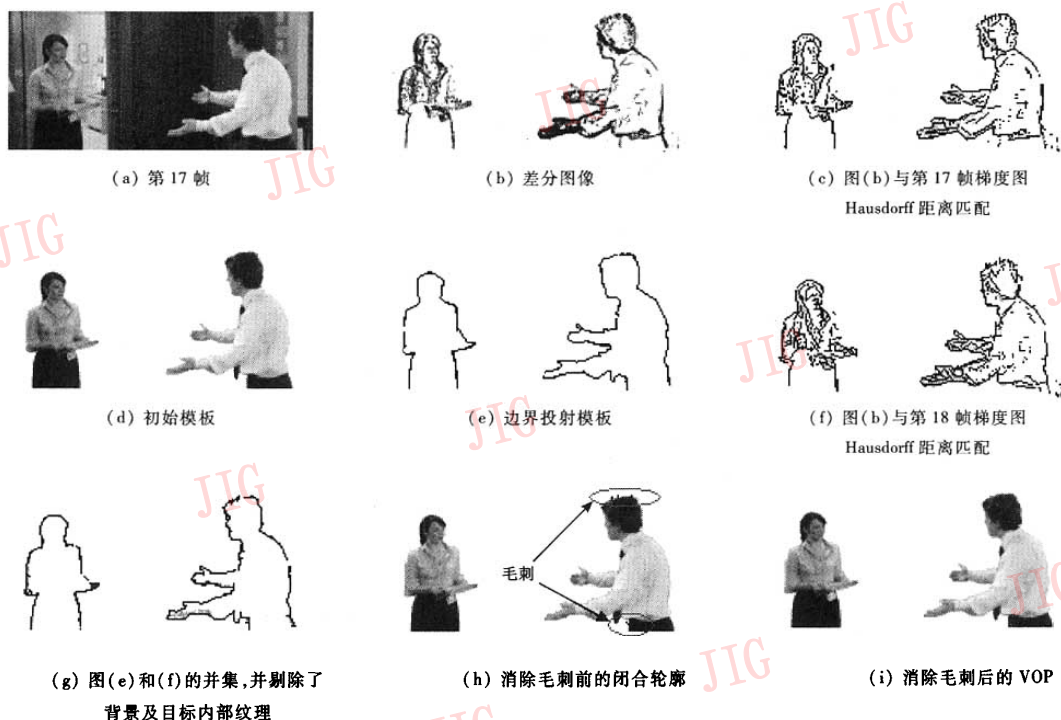


图 7 电影《真爱至上》序列提取 VOP 的过程

Fig. 7 The process of obtaining VOP in the film "Love Actually"

补缺失的左臂轮廓线,从图 7(g)可见,目标轮廓线虽未闭合,但缺口很小,这为在  $N \times N$  邻域内自动闭合轮廓线提供了必要条件。经过梯度图像和差分图像的 Hausdorff 距离匹配后,仍保留了显露出的背景和内部纹理,如图 7(f)所示,男士左臂上方有许多细碎的背景纹理。采用 3.3 节的方法,成功地消除了它们,如图 7(g)所示。闭合轮廓线后,发现男士头顶及裤子上有毛刺,这是因为它们与真正的轮廓点相连通,采用 3.3 节的方法及滤除噪声的方法,均无法消除它们。对此,采用 3.4 节的第 5 步成功地消除了这些毛刺,如图 7(i)所示。

这些场景中均有两个运动目标,从实验结果可以看到,本算法成功地提取了这两个目标的精确轮廓。

## 5 结论

把帧间运动信息和帧内的梯度信息结合起来,提出了一种跟踪运动目标并获得精确 VOP 的简单快速算法。该算法通过边界投射模板的补偿作用,形成下一帧的完整边缘轮廓,以达到在复杂背景下

有效跟踪多个非刚体运动目标的目的,较好地解决了被遮挡背景的显露问题,避免了将复杂背景并入运动目标的错误。本算法还存在着一些问题没有解决,例如,该算法无法处理两个分离目标在运动过程中相互遮挡以及目标的合并与分离的情况。此问题是今后进一步研究的内容。

## 参考文献 (References)

- 1 JI Bai-yang, CHEN Chun, QIAN Ying. The development of video segmentation technology [J]. Journal of Computer Research and Development, 2001, 38(1): 36 ~ 42. [季白杨,陈纯,钱英. 视频分割技术的发展[J]. 计算机研究与发展, 2001, 38(1): 36 ~ 42.]
- 2 Fan Jianping, David K Y Yau. Automatic image segmentation by integrating color-edge extraction and seeded region growing [J]. IEEE Transactions on Image Processing, 2001, 10(10): 1454 ~ 1466.
- 3 JIA Zhen-tang, LI Sheng-ping, HE Gui-ming, et al. A video object segment algorithm based on moving edge detection [J]. Journal of Computer Research and Development, 2003, 40(5): 684 ~ 689. [贾振堂,李生平,贺贵明等. 一种基于运动边缘检测的视频对象分割新算法[J]. 计算机研究与发展, 2003, 40(5): 684 ~ 689.]
- 4 Milan sonka, Vaclav Hlavac, Roger Boyle. Image Processing, Analysis, and Machine Vision the second edition [M], Beijing: Posts and Telecom Press, 2002: 374 ~ 380.

- 5 Thomas Meier, King N Ngan. Automatic segmentation of moving objects for video object plane generation [J]. IEEE Transactions on Circuits and Systems for Video Technology, 1998, 8(5): 525 ~ 538.
- 6 Thomas Meier, King N Ngan. Video segmentation for content-based coding [J]. IEEE Transactions on Circuits and Systems for Video Technology, 1999, 9(8): 1190 ~ 1203.
- 7 Wang Fangshi, Xu De. A semiautomatic segmentation algorithm of multiple nonrigid moving objects in clutter background [A]. In: Proceedings of International Conference on Signal Processing (ICSP' 2004) [C], Beijing, 2004: 785 ~ 788.
- 8 YU Tian-li, ZHANG Yu-jin. Local motion information-based video retrieval method [J]. Journal of Tsinghua University (Science & Technology), 2002, 42(7): 925 ~ 928. [俞天力, 章毓晋. 一种基于局部运动特征的视频检索方法 [J]. 清华大学学报(自然科学版), 2002, 42(7): 925 ~ 928.]