

# 一种基于统计推断的自动视频对象分割算法

赵明 李娜 陈纯

(浙江大学计算机科学与技术学院 杭州 310027)

(浙江大学虹软多媒体研究中心 杭州 310027)

**摘要** 在新一代 MPEG-4 视频编码标准中,为了支持面向对象编码和实现基于内容的应用,视频对象(VO)的自动分割成为关键技术之一。减背景法是视频对象自动分割的基本方法,但是不同的环境光照条件常常给视频对象的分割带来困难。本文提出一种基于统计推断的减背景法,首先建立背景统计模型,然后对后续帧进行假设检验,从而分割出视频对象。同时,本文算法采用 HSV 颜色空间,并且通过对背景统计模型中各颜色分量的有效分析和区别使用,能够很好地适应不同的环境光照条件。实验表明,本文算法能够在各种光照环境下自动地实现视频对象的准确分割。

**关键词** 视频分割,面向对象编码,MPEG-4

中图法分类号 TP391.4

## 1 引言

随着视频编码技术的发展,面向对象编码日益成为研究热点。在新一代视频编码标准 MPEG-4 中,视频序列中的图像被表示为视频对象<sup>[1]</sup>(VO)的集合,这不仅大大提高编码效率,而且使基于内容的应用成为可能。然而,MPEG-4 却假设在编码之前,视频序列已经被表示为视频对象。所以,视频对象的获得(即视频对象分割)对于 MPEG-4 以及其相关特性的实现至关重要。

视频分割技术通常分为自动分割技术和半自动分割技术。从背景静止的视频流中提取视频对象是自动分割技术的一种有效方法,其基本方法是变化检测法<sup>[2]</sup>。变化检测算法的基本思想是根据两帧图像中的亮度、颜色或纹理变化来检测并分割视频对象。变化检测法有相邻帧相减法和减背景法两种方案。相邻帧相减法虽然可以降低环境光线变化的影响,但是由于视频对象运动的多样性,使得利用相邻帧相减往往得不到视频对象的完整轮廓。虽然文献<sup>[3]</sup>采用 Hausdorf 距离进行轮廓匹配,并利用边缘等信息来恢复运动视频对象的完整轮廓,但是它不仅计算量大而且精度低。减背景法可以很好地解决上述问题,但是要满足一个条件:视频流中必须有一定数量的背景帧(只有静止背景、不含视频对象),在一定的场合下,这种条件容易满足。由于减背景法的分割效果好,速度快,所以得到了广泛的关注<sup>[4][5][6]</sup>。

Horprasert<sup>[4]</sup>等人基于 RGB 颜色空间,提出一种将色度与亮度分离的模型。在建立背景模型时计算各像素的颜色均值,然后将待检测图像中每个像素的颜色值与对应像素的颜色均值的差异分解为亮度偏差和色度偏差,并利用它们来检测视频对象。这种分割算法能够在一定程度上解决光线变化和阴影问题,但是只能抑制微小的光线变化,并且要求环境背景有良好的光照条件,当环境光线较暗或变化较大时,效果非常差。MIT 的 Pfinder<sup>[5]</sup>系统采用基于 YUV 颜色空间的检测算法,取得了较好的分割效果。但该算法只能适应光线的缓慢变化,环境光线突然改变时分割效果较差。Francois<sup>[6]</sup>等人采用基于 HSV 颜色空间的变化检测算法,分割效果较好。但是该算法的缺点在于只对当前帧进行比较,所以有较多噪声,分割不精确,同时没有对 HSV 颜色空间中的各颜色分量进行合理分析和区别使用,算法强壮性较差。

本文提出一种基于统计推断的减背景法,自动地实现视频对象的准确分割。本文算法与其他自动视频对象分割算法相比有以下特点:(1)利用统计推断中的假设检验分割视频对象,使算法有较少的噪声和较高的准确性。(2)基于 HSV 颜色空间进行分析处理,将亮度信息和颜色信息分开,提高了算法的稳定性。(3)根据各颜色分量的不同特性和每个像素各颜色分量的不同统计特征进行区别使用,使算法具有很强的鲁棒性。

## 2 视频对象分割算法

本文提出的视频对象分割算法主要由以下几个部分组成:将每个像素的颜色从 RGB 颜色空间转

换到 HSV 颜色空间；利用背景帧建立背景统计模型；对后续帧进行假设检验得出视频对象的大体轮廓；通过后处理得到视频对象的精确轮廓。下面对它们进行分别讨论。

## 2.1 转换到 HSV 颜色空间

大多数视频捕捉设备采用 RGB 颜色空间。但是由于 R、G、B 三个颜色分量存在相关性，所以在面向彩色处理时通常不直接使用 RGB 颜色空间，而采用其变形或其他颜色空间。HSV 颜色空间把颜色表示成色度 (H)、饱和度 (S) 和亮度 (V) 三个分量，它们独立感知颜色变化，相互独立，互不影响，从而能够提高算法的稳定性。虽然从 RGB 到 HSV 的转换耗费一些时间，但相对于当前 PC 机的处理速度是完全可以接受的，所以本文采用了 HSV 颜色空间。

## 2.2 建立背景模型

本文采用统计模型对静态背景中每个像素的每个颜色  $K \in \{H, S, V\}$  都进行建模。由于摄像头存在均值为 0 的高斯白噪声，所以可以用高斯分布模型表示静态背景中每个像素在一段时间内的颜色分布。

取视频流中不包含视频对象的前  $N$  帧（只有静态背景）对背景进行统计建模，即对前  $N$  帧中的每一点  $X$ ，求出其均值  $\mathbf{m}_{XK}$  和标准差  $\mathbf{s}_{XK}$  ( $K \in \{H, S, V\}$ )，它们表示背景在正常情况下的变化特征。当视频对象进入后，如果某点的变化超出了正常变化范围，则有理由认为此点不是背景，而是视频对象。在前  $N$  帧中，像素  $X$  的颜色分量  $K \in \{H, S, V\}$  的取值集合为  $B^{(N)}_{XK} = \{K_i | i=1, \dots, N\}$  ( $K \in \{H, S, V\}$ )，它是满足正态分布的一个样本，其样本均值和样本标准差分别为  $\bar{B}^{(N)}_{XK} = \frac{1}{N} \sum_{i=1}^N K_i$  ( $K \in \{H, S, V\}$ )，

$$S^{(N)}_{XK} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (K_i - \bar{B}^{(N)}_{XK})^2} \quad (K \in \{H, S, V\})。$$

由于  $\bar{B}^{(N)}_{XK}$  和  $S^{(N)}_{XK}$  分别是总体均值  $\mathbf{m}_{XK}$  和总体标准差  $\mathbf{s}_{XK}$  的

无偏估计,所以可用它们表示  $\mathbf{m}_{XK}$  和  $\mathbf{s}_{XK}$ ，即  $\mathbf{m}_{XK} = \bar{B}^{(N)}_{XK}$ ， $\mathbf{s}_{XK} = S^{(N)}_{XK}$ 。

## 2.3 统计推断

建立背景模型之后，就可以根据已有的背景模型来推断后来的数据是背景还是视频对象。由于视频帧中每个像素  $X$  的各颜色分量  $K \in \{H, S, V\}$  相互独立，如果后来的数据是背景，那么每个颜色分量都不能有太大偏差，所以必须对它们分别进行统计推断。本文采用假设检验进行统计推断，由于通过背景建模，背景在正常情况下的均值  $\mathbf{m}_{XK}$  和标准差  $\mathbf{s}_{XK}$  已知，所以统计推断的任务就是检验后来数据的均值是否与背景均值相符。如果不相符，则表示有视频对象出现，引起了背景的异常变化。

用  $\bar{\mathbf{m}}^{(m)}_{XK}$  ( $K \in \{H, S, V\}$ ) 表示后继帧中相邻  $m$  帧像素点  $X$  的颜色分量  $K$  的均值，对它的假设检验为：

$$H_0: \bar{\mathbf{m}}^{(m)}_{XK} = \mathbf{m}_{XK}, \quad H_1: \bar{\mathbf{m}}^{(m)}_{XK} \neq \mathbf{m}_{XK}。$$

其检验统计量为  $z = \frac{\bar{\mathbf{m}}^{(m)}_{XK} - \mathbf{m}_{XK}}{\mathbf{s}_{XK}/\sqrt{m}}$ ，设在  $H_0$  为真时拒绝  $H_0$  的最大概率

为  $\alpha_k$  (也称显著性水平)，则可以求出其拒绝域为<sup>[7]</sup>  $|z| \geq z_{\alpha_k/2}$  即  $|\bar{\mathbf{m}}^{(m)}_{XK} - \mathbf{m}_{XK}| \geq z_{\alpha_k/2} \mathbf{s}_{XK}/\sqrt{m}$ ，此时拒绝

$H_0$ ，接受  $H_1$ ，也就是  $\bar{\mathbf{m}}^{(m)}_{XK} \neq \mathbf{m}_{XK}$ ，因此有理由认为当前像素的值不是背景正常情况下的取值，所以此

时该像素属于视频对象。当  $|z| < z_{\alpha_k/2}$ ，即当  $|\bar{\mathbf{m}}^{(m)}_{XK} - \mathbf{m}_{XK}| < z_{\alpha_k/2} \mathbf{s}_{XK}/\sqrt{m}$  时，没有理由拒绝  $H_0$ ，所以认为此时该像素属于背景。

由于本文在假设检验时,对后继帧中相邻 $m$ 帧的均值  $\bar{\mathbf{m}}_{XK}^{(m)}$  进行检验,所以具有抑制噪声的功能,使分割结果更准确。但是  $m$  也不能太大,否则会平滑视频对象引起的变化,一般情况下取 3 至 7。

## 2.4 视频对象分割

由于视频流的不稳定性和多样性,所以在采用统计推断分割视频对象之前,必须考虑到视频流中的特殊情况,才能使算法具有很好的鲁棒性。下面先解决将这些特殊情况所产生的问题,然后给出本文的视频对象分割算法。

### 2.4.1 阴影检测

由于光源的原因,视频对象可能在背景中留下阴影。本文在分割视频对象时,认为阴影不属于视频对象,因为它是由环境的特殊光照产生的。但是,其颜色特征又与背景有较大的区别,很容易被检测成视频对象,故必须对阴影进行特殊的分析和处理。

从颜色特性分析知道,视频对象在背景中留下的阴影区域与对应的原背景区域只是亮度有较大不同,而颜色的其他信息(色度和饱和度)却几乎没有变化。由于本文采用了 HSV 颜色空间,其亮度信息(V)与色度(H)和饱和度(S)是分离的,所以用来检测阴影不仅相当方便,而且不会受其他因素的影响,算法稳定。采用 HSV 颜色空间进行阴影检测时,阴影区域必须满足以下条件:

$$\begin{cases} \bar{\mathbf{m}}_{XV}^{(m)} - \mathbf{m}_{XV} \leq -z_{\alpha_V/2} \mathbf{S}_{XV} / \sqrt{m} \\ |\bar{\mathbf{m}}_{XH}^{(m)} - \mathbf{m}_{XH}| < z_{\alpha_H/2} \mathbf{S}_{XH} / \sqrt{m} \\ |\bar{\mathbf{m}}_{XS}^{(m)} - \mathbf{m}_{XS}| < z_{\alpha_S/2} \mathbf{S}_{XS} / \sqrt{m} \end{cases}$$

### 2.4.2 光线变化与背景更新

在视频流中,环境的全局光线经常有缓慢的变化,同时当视频对象在场景中运动时,也可能改变背景环境的局部光照。如果不对原来建立的背景模型进行更新就容易将背景检测成视频对象,因此必须动态地更新背景模型,即让背景模型符合最近的环境光线变化。文献<sup>[5]</sup>使用基于概率的自适应模型进行背景更新,文献<sup>[8]</sup>用带有预测的 Kalman 滤波器来适应变化,文献<sup>[6]</sup>用线性迭代的方法对背景区域的像素进行更新,即  $\mathbf{m} = (1-a)\mathbf{m} + a\mathbf{m}$ 。然而它们对于静态背景并不是非常适宜,因为首先背景本身没有变化,只是背景的光照变化,其次光照变化是暂时的,特别是阴影,不会影响所有后续的视频帧,变化结束后,背景又恢复原来的状态。而文献<sup>[5][6][8]</sup>却在背景模型中保留先前所有发生过的变化,反而会让初始背景模型发生较大的失真,使背景模型没有起到应有的作用。鉴于此,本文提出了如下的背景更新算法。

用  $\mathbf{m}_{XK}^N$  表示不包含视频对象的前  $N$  帧(只有静态背景)建立的初始背景模型,在建立背景模型阶段刚结束时  $\mathbf{m}_{XK}^N = \mathbf{m}_{XK}$ 。对于后续帧中的每一点  $X$ ,在它不是前景并且不是阴影的条件下,如果

$|\bar{\mathbf{m}}_{XK}^{(m)} - \mathbf{m}_{XK}^N| \geq T_{UPDATE}$ ,那么  $\mathbf{m}_{XK} = (F_{XK} \mathbf{m}_{XK} + \mathbf{b}_{XK} \bar{\mathbf{m}}_{XK}^{(m)}) / (F_{XK} + \mathbf{b}_{XK})$ ,其中  $\mathbf{b}_{XK} = \mathbf{g}_{XK} |\bar{\mathbf{m}}_{XK}^{(m)} - \mathbf{m}_{XK}^N|$ ,表示更新速度,  $F_{XK}$  表示已经处理过的帧数,包括建模用的前  $N$  帧,在更新  $\mathbf{m}_{XK}$  之后  $F_{XK} = F_{XK} + 1$ ,  $T_{UPDATE}$  表示更新

阈值,只有光线变化大于它时才进行背景更新;如果  $|\bar{\mathbf{m}}_{XK}^{(m)} - \mathbf{m}_{XK}^N| < T_{UPDATE}$ ,那么  $\mathbf{m}_{XK} = \mathbf{m}_{XK}^N$  并且  $F_{XK} = N$ ,因为此时光线又恢复初始的状态,所以必须去处光线变化过程中对背景模型产生的影响。

本文提出的背景更新算法充分考虑了光线变化的暂时性和初始建立的背景模型,所以收到了很好的效果。

### 2.4.3 颜色分量的变化特征

尽管 HSV 空间符合人眼的视觉特性,并且将各个分量独立地放开,但是其各个分量并非在各种

情况下都非常稳定，在较暗区域，S 和 H 变化很大；在高反射区域，H 变化较大，如图（1）所示。

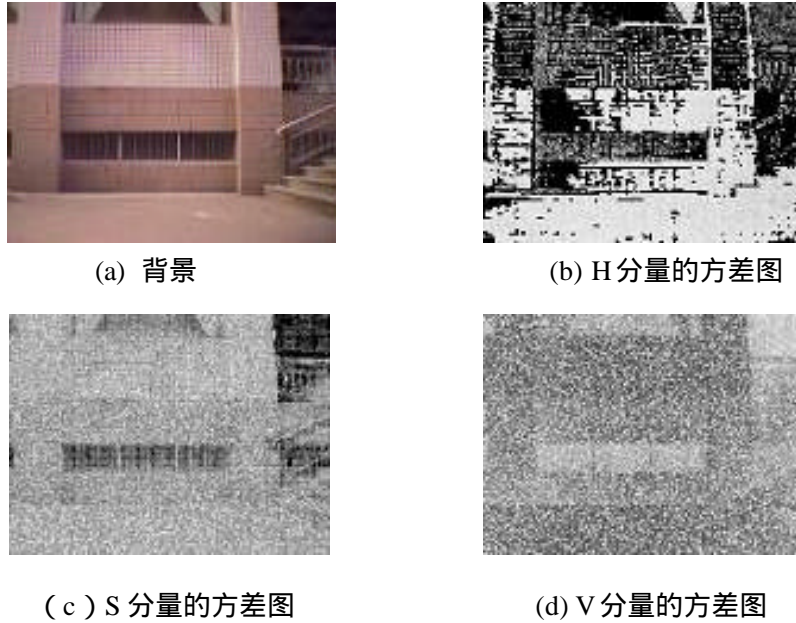


图 1 静态背景中 HSV 各分量的方差（灰度值越低表示方差越大）

Fig.1 the variance of the color components of the HSV color space (the darker, the larger)

由图 1 可见，由于背景中各区域的亮度和反射特性的不同，H、S、V 各分量的稳定性在各个区域有很大差别。总体上说，H 分量最不稳定，S 分量次之，V 分量最稳定。对于某个分量变化剧烈的像素点，如果利用背景模型对此颜色分量也进行统计推断，则分割效果会很差，因为此颜色分量已经偏离了假设的正态分布，统计推断不能适用。但是 V 分量一般情况下总是可靠的，故只须检查 H、S 颜色分量的标准差，如果大于给定阈值，即如果  $s_{xk} \geq T_k$ ，则不用该分量来进行统计推断。

#### 2.4.4 基于统计推断的视频对象分割算法

首先取视频流中不包含视频对象的前  $N$  帧（只有静态背景）在 HSV 颜色空间建立背景统计建模，即对前  $N$  帧中的每一点  $X$ ，先将其颜色从 RGB 空间转换到 HSV 空间，并求出其均值  $\mathbf{m}_{xk}$  和标准差  $s_{xk}$  ( $K \in \{H, S, V\}$ )。

然后对后继每一帧中的像素点  $X$ ，先将其颜色从 RGB 空间转换到 HSV 空间，然后利用以这一帧为中心的  $m$  帧求出  $\bar{\mathbf{m}}_{xk}^{(m)}$  ( $K \in \{H, S, V\}$ )，并执行以下算法：

(1) 如果  $\bar{\mathbf{m}}_{xv}^{(m)} - \mathbf{m}_{xv} \geq z_{\alpha_v/2} s_{xv} / \sqrt{m}$ ，则  $X$  属于视频对象，否则执行下一步

(2) 如果  $\bar{\mathbf{m}}_{xv}^{(m)} - \mathbf{m}_{xv} \leq -z_{\alpha_v/2} s_{xv} / \sqrt{m}$ ，则转 (5)，否则执行下一步

(3) 如果  $|\bar{\mathbf{m}}_{xs}^{(m)} - \mathbf{m}_{xs}| \geq z_{\alpha_s/2} s_{xs} / \sqrt{m}$  并且  $s_{xs} \leq T_s$ ，则  $X$  属于视频对象，否则执行下一步

(4) 如果  $|\bar{\mathbf{m}}_{xh}^{(m)} - \mathbf{m}_{xh}| \geq z_{\alpha_h/2} s_{xh} / \sqrt{m}$  并且  $s_{xh} \leq T_h$ ，则  $X$  属于视频对象，否则属于背景，并对  $X$  进行背景更新。

(5) 如果  $|\bar{\mathbf{m}}_{xs}^{(m)} - \mathbf{m}_{xs}| < z_{\alpha_s/2} s_{xs} / \sqrt{m}$  并且  $|\bar{\mathbf{m}}_{xh}^{(m)} - \mathbf{m}_{xh}| < z_{\alpha_h/2} s_{xh} / \sqrt{m}$ ，则  $X$  属于背景中的阴影，否则  $X$  属于视频对象。

分割算法的结果用一张二值图表示，0 表示背景，1 表示视频对象。



## 2.5 后处理

### 2.5.1 噪声过滤

由于本文的视频对象分割算法是基于像素的，因此在算法结果二值图中会产生许多离散点，它们可能是视频对象，也可能是噪声点。因此，一方面需要将属于视频对象的离散点连接起来，另一方面又要去处噪声点。形态学滤波对于小噪声比较有效，对于大噪声效果不理想。由于基于像素法的视频分割可能产生大片的噪声，所以本文采用另一种卷积核来过滤噪声，其方法是对每一个像素，判断以该像素坐标为中心的一定大小的窗口内值为 1 的像素个数，如果大于指定的阈值，则将该像素的值置为 1，即认为属于视频对象，否则置为 0，认为是噪声。可根据噪声大小调整窗口尺寸，考虑到计算速度，在实际应用中的卷积核为如图 2 所示，即只计算窗口外边缘上值为 1 的像素的个数，在噪声较大的情况下，采用这种卷积核滤波的效果好于形态学滤波。

```
1 1 1 1 1
1 0 0 0 1
1 0 0 0 1
1 0 0 0 1
1 1 1 1 1
```

图 2

噪声过滤卷积核

Fig.2 convolving  
kernel for filtering  
noise

### 2.5.2 区域填充

在过滤噪声点或小噪声块后，还可能存在面积较大的噪声块，同时视频对象内可能会出现空洞。对于前者，本文同样采用基于链码的一遍扫描快速算法<sup>[9]</sup>，标记出各连通块，并将包含像素数目小于指定阈值的区域视为噪声块删去。对于后者，只需要对选择出来的连接块中的空洞做填充处理即可。本文采用的方法是先在二值图像的四周扩展一圈灰度值为 0 的像素，然后对这个二值图中灰度值为 0 的像素进行种子填充，种子选在扩展后的图像的原点。能够填充的区域便是背景区域，不能填充的像素用灰度值 1 填充，最后灰度值为 1 的像素组成的区域便是视频对象所在的区域。

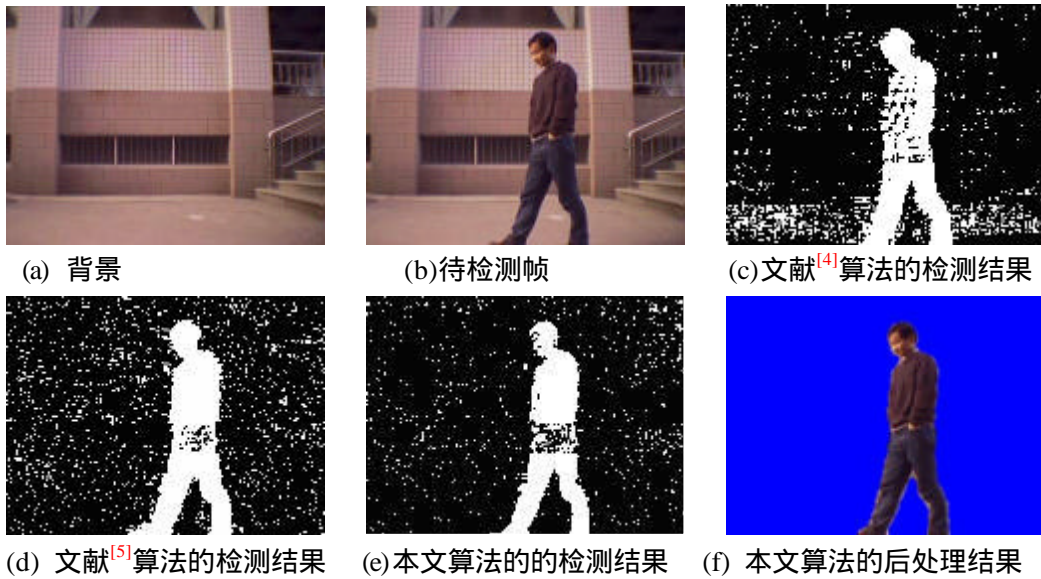


图 3 室外视频流分割结果

Fig.3 the segmentation results for outdoor video stream

## 3 实验结果

本文算法在 PIII 500 机器上，对  $320 \times 240$  分辨率的视频处理速度达到 7 帧/秒。下面首先将本文算法对视频对象的分割效果与文献<sup>[4] [5]</sup>算法的分割效果进行对比。为了更好地说明问题，测试视频流选用不同质量的室内和室外视频流。

第一个是室外视频流（如图 3 所示），光照条件好，光线均匀且有足够的亮度，视频对象进入背

景不会引起背景光线的明显变化，且不会在背景中产生阴影。文献<sup>[4]</sup>的算法在图像的底部有大量的错检点，文献<sup>[5]</sup>的算法噪声点较均匀，但噪声点比本文的算法要密集，特别是在对象周围有较多噪声。本文算法的噪声比较分散，因此可以通过后处理来有效地消除，分割效果较好。

第二个是室内视频流（如图 4 所示），光照条件差，光线不均匀且亮度低，视频对象进入背景后引起背景光线明显变化，同时在背景中产生阴影。此时，文献<sup>[4]</sup>算法把大片背景错检为视频对象，文献<sup>[5]</sup>算法错检点虽然更分散，但是却形成很多错检区域。本文算法的错检点少且分散，可通过后处理来有效地消除，因此能够很好地分割出视频对象。

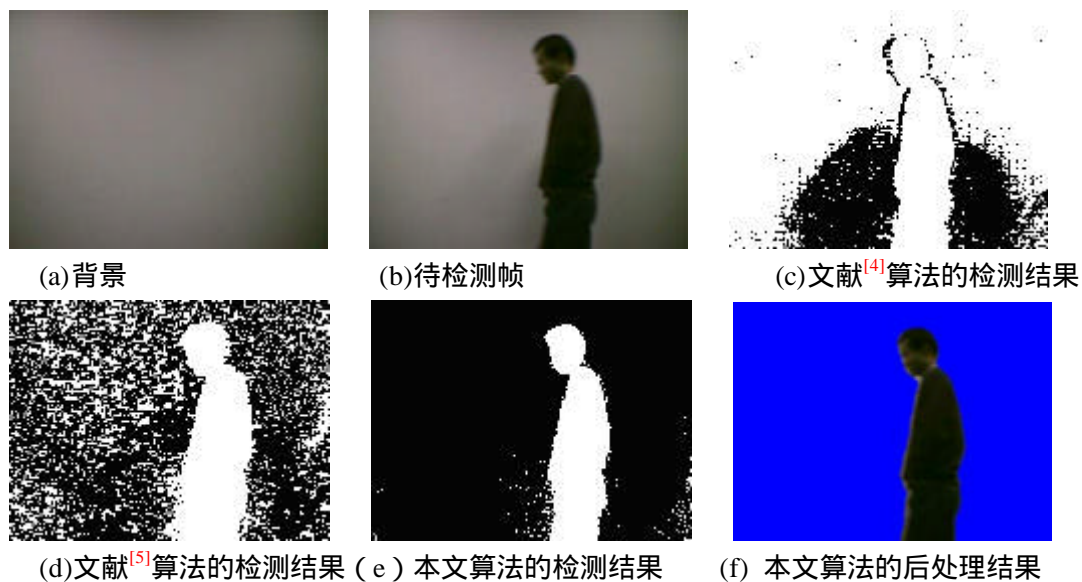


图 4 室内视频流分割结果

Fig.4 the segmentation results for outdoor video stream

由以上两个测试视频序列可以看出，本文算法的视频对象分割精确度要高于文献<sup>[4]</sup> <sup>[5]</sup>的算法。文献<sup>[4]</sup>的算法在照明充分、摄像头离场景距离较远的场合方能取得较好的效果。文献<sup>[5]</sup>的算法受背景的复杂性影响较大，在光线变化较大时效果也不好。

最后给出本文算法对 MPEG-4 测试序列 Hall Monitor 的视频对象分割的结果，如图 5 所示。由分割结果可以看出，本文算法能够较好地实现视频对象的自动分割。

## 4 结论

本文提出一种基于统计推断的自动视频分割算法。算法利用统计推断中的假设检验分割视频对象，使算法有较少的噪声和较高的准确性。同时采用了 HSV 颜色空间，并且在此空间上进行了阴影检测和背景更新，特别地，根据 HSV 空间中各颜色分量的不同特性和每个像素各颜色分量的不同统计特征进行区别使用，使算法具有很强的鲁棒性，能够很好地适应不同的环境光照条件。实验表明，本文算法能够在各种光照环境下自动地实现视频对象的准确分割，可用以实现面向对象编码、基于内容的多媒体应用、视频监控、视频会议等。本文下一步的工作是进一步提高速度，并将算法推广到非静止背景的场所。

### 参考文献

- [1] MPEG Video and SNHC Groups. "Committee draft of MPEG-4, part 2, 14496-2"[R] Technical Report ISO/IEC JTC/SC29/WG11/N1902, Fribourg, Switzerland: ISO/IEC, October 1997.
- [2] Ji Baiyang, Chen Chun, Qian Ying. The development of video segmentation technology[J]. Journal of computer research and development. 2001, 38(1):36-42 (in Chinese). 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.)

- [3] M.Thomas, N.N.King. automatic segmentation of moving objects for video object plane generation[J]. IEEE Trans on Circuits and System for Video Technology, 1998, 8(5): 525-538
- [4] T. Horprasert, D. Harwood, and L.S. Davis, a statistical approach for real-time robust background subtraction and shadow detection[A]. IEEE international conference on computer vision FRAME-RATE Workshop, <http://www.eecs.lehigh.edu/FRAME/Horprasert/index.html>, 1999.
- [5] C.Wren, etc, "PFinder:Real-Time Tracking of the Human Body"[J], IEEE Transaction on Pattern Analysis and Machine Intelligence, 1997,19(7):780-785.
- [6] A. Francois and G. Medioni, "Adaptive Color Background Modeling for Real-Time Segmentation of Video Streams"[A]. In: Proceedings of International on Imaging Science, Systems, and Technology. Las Vegas, NA, 1999:227-232.
- [7] Sheng Zhou, Xie Shiqian, Pan Chengyi. The probability and statistics[M]. the second edition. Beijing: The higher education press.1989:205-206 (in Chinese). (盛骤,谢式千,潘承毅. 概率论与数理统计(第二版)[M].北京:高等教育出版社. 1989:205-206.)
- [8] C.Ridder, O.Munkelt, H.Kirchner, "Adaptive background estimation and foreground detection using Kalman-filtering"[A], In: Proceedings of International Conference on recent Advances in Mechatronics. Istanbul, Turkey, 1995:193-199.
- [9] Jae-Chang Shim. Chitra Dorai. a generalized region labeling algorithm for image coding, restoration, and segmentation[A]. In: Proceedings of international conference on image processing. Kobe, Japan, 1999:46-50.

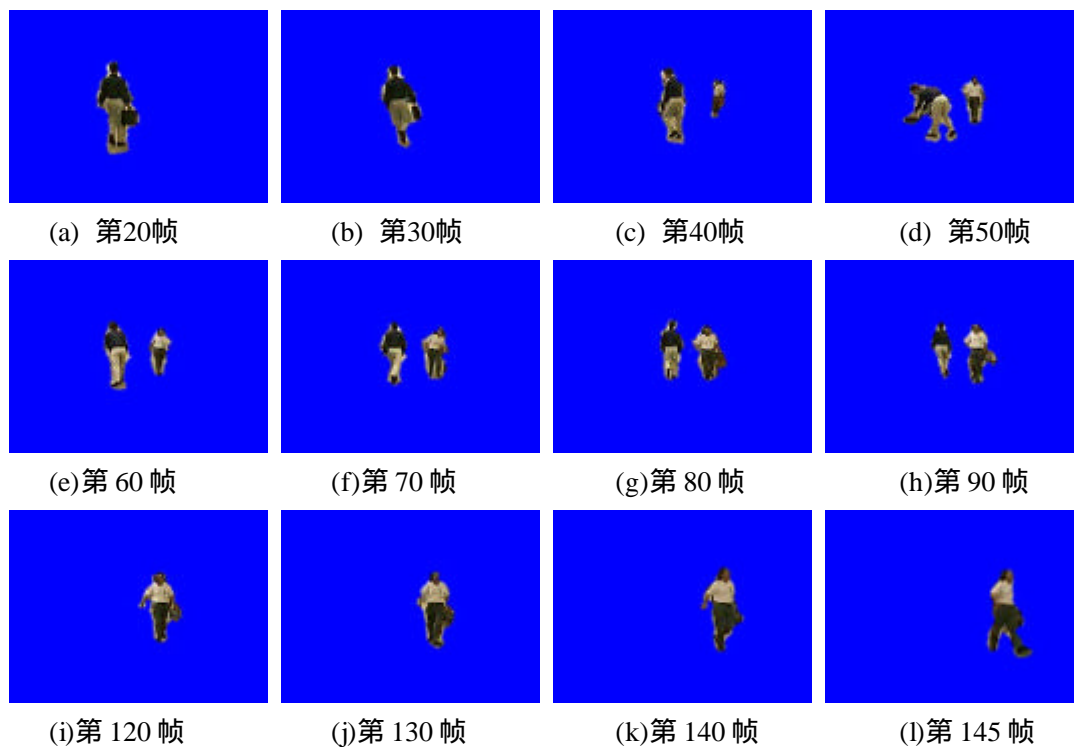


图5 对视频序列 Hall Monitor 的分割结果  
Fig.5 the segmentation results for Hall Monitor

# A Statistical Inference Based Automatic Video Object Segmentation Algorithm

ZHAO Ming LI Na CHEN Chun

(Department of Computer Science and Engineering, Zhejiang University, hangzhou, 310027)

( Zhejiang University Arcsoft Multimedia Research Center , hangzhou, 310027 )

**Abstract** In the new MPEG-4 video coding standard, the automatic video object segmentation plays a key role in supporting object-oriented coding and enabling content-based functionalities. Background subtraction is one of the basic automatic video object segmentation method. But various environmental illumination conditions often make it hard to work. A statistical inference based background subtraction method is presented in this paper. A statistical background model is firstly setup in this algorithm. Then the hypothesis testing is applied to the following frames to segment the video objects. The HSV color model is also used and its color components are efficiently analyzed and treated separately so that the proposed algorithm can adapt to different environmental illumination conditions. The experimental results show this algorithm can automatically segment video objects accurately in various illuminating environment.

**Keywords** video segmentation, object based coding, MPEG-4