

视频监控前景图像估计的盲源提取方法*

王群¹, 薛瑞¹, 孙振江²

(1. 北京航空航天大学 电子与信息工程学院, 北京 100191; 2. 国防科技大学 教研保障中心, 湖南 长沙 410073)

摘要:在视频图像运动检测的背景消减方法中,场景图像或帧可建模为前景图像和背景图像的叠加或线性混合。然而,实际中图像的背景和前景往往相关,常用的主成分分析和独立分量分析等方法难以实现准确提取。为此,将视频图像的前景提取建模为盲源提取问题,提出了一种基于均方交叉预测误差的盲源提取方法,可以从相关的源视频图像中提取期望的前景图像,并将该方法扩展应用于基于基本模型和特征背景模型背景消减方案中。基于人工和实际视频的实验验证了盲源提取背景消减方法的可行性和有效性。

关键词:运动检测;背景消除;前景分离;盲源提取;均方交叉预测误差

中图分类号:TN95 **文献标志码:**A **文章编号:**1001-2486(2019)01-130-12

Foreground estimation in video surveillance by blind source extraction

WANG Qun¹, XUE Rui¹, SUN Zhenjiang²

(1. School of Electronics and Information Engineering, Beihang University, Beijing 100191, China;

2. Teaching and Researching Supporting Center, National University of Defense Technology, Changsha 410073, China)

Abstract: In video surveillance, one scene image/frame can be modeled as a superimposition or linear mixture of foreground visual contents and background contents. In the real world, however, the background and foreground are correlated to each other. Therefore, the foreground extraction cannot be well solved by the PCA (principle component analysis) and the ICA (independent component analysis) algorithms. The foreground extraction was modeled as a BSE (blind source extraction) problem. The MSCPE (mean square cross prediction error), one solution of BSE, was generalized to extract desired source signal which was correlated with other source signals. Then MSCPE BSE method was applied to the background subtraction schemes by using the basic model and eigen backgrounds method. Experimental results on artificial video shows the feasibility of MSCPE, and the real-world video experiments demonstrate its effectiveness.

Keywords: motion detection; background subtraction; foreground segmentation; blind source extraction; mean square cross prediction error

运动检测是从图像序列中检测运动目标,在视频监控的目标跟踪、活动识别和行为理解中具有重要作用^[1-16]。运动检测的基本原理是从视频图像的背景中提取运动目标,也称为前景图像序列。在大多数视频监控应用中,如交通或临床监视,摄像头是固定的。因此,在静态摄像机下检测图像序列中,运动物体的运动检测方法受到广泛关注,主要包括背景消减法和时间差分法^[16]。由于时间差分法对于环境噪声较为敏感,仅适用于部分室内场景,因此本文重点研究背景消减法。

传统背景消减法通过评估当前场景图像与参考背景图像像素特征的差异来提取前景图像。该方法易于实现,但对光照变化非常敏感。为了适应环境(如光照)的变化,学者们提出多种改进方法来自适应地更新参考背景,主要包括两类:第一

类是改进方法关注如何用合适的概率分布对背景像素进行建模^[3-13];第二类是改进方法关注如何构建不包含移动对象的参考背景图像^[14-16]。

在第一类改进方法中,背景中的每个像素都被建模为服从特定分布(如高斯分布^[3-4]和混合高斯分布^[5-9])。文献[10]提出了用非参数核密度估计来评估背景像素的概率,文献[11-12]对其进行了改进,以减小计算量。文献[13]提出了视频背景提取算法,为每个像素存储一组过去在相同位置或邻近区域获取的值,通过将该集合与当前像素值进行比较,以确定该像素是否属于背景。然而,在实际应用中,针对较快变化的背景,上述方法的背景更新过程占用了大量的计算时间。随着图像尺寸的增加,第二类改进方法受到越来越多的关注。

* 收稿日期:2018-01-22

基金项目:国家自然科学基金资助项目(91438207)

作者简介:王群(1982—),女,北京人,博士研究生,E-mail:wangqun_boco@hotmail.com;

薛瑞(通信作者),男,副教授,博士,硕士生导师,E-mail:xuerui@buaa.edu.cn

第二类改进方法将运动目标视为前景图像,视频监视图像视为背景图像和前景图像的叠加或线性混合,将前景对象作为当前帧和参考背景图像之间的差异进行检测^[14],因此计算速度快。学者们分别提出了基于主成分分析(Principal Component Analysis, PCA)^[15]、独立成分分析(Independent Component Analysis, ICA)的方法^[16],在图像处理中获得了广泛应用^[17-19]。然而,实际上视频的背景图像和前景图像通常是相关的。基于PCA的方法仅能提取方差最大的主成分,而主成分可能不包含前景图像信息。基于ICA的方法要求背景图像和前景图像独立,也无法进行很好的提取。因此,如何从相关的源图像中提取期望的前景图像成为一个难题。

在信号处理领域,盲源提取(Blind Source Extraction, BSE)是一种利用先验特性提取期望信号的重要手段^[20-21],与PCA、ICA等盲源分离(Blind Source Separation, BSS)方法相比,BSE精度更高、鲁棒性更好。BSE包括使用高阶统计量(Higher Order Statistics, HOS)^[20]和二阶统计量(Second Order Statistic, SOS)^[22-26]两类方法。其中,基于线性预测器或自回归(AutoRegressive, AR)模型^[22-23]SOS的BSE方法计算量小,适用于对实时性要求较高的应用。Wang等证明了当任一期望信号的AR参数已知时,如果源信号相互独立,基于均方交叉预测误差(Mean Square Cross Prediction Error, MSCPE)的BSE算法可以提取期望信号^[22-23]。MSCPE能满足特定AR模型参数的要求,并能根据给定的AR参数提取任意期望信号。

基于上述研究,本文将MSCPE从独立源信号中期望信号的提取推广到相关源信号中期望信号的提取,提出一种背景消减的盲源提取方法,可以在前景图像与背景图像相关的条件下,直接从混合观测矩阵中提取前景图像,计算速度快,且可以适应不同光照变化。

本文首先给出了视频监视图像中运动检测的线性混合模型,然后给出了基于MSCPE从相关源中提取期望信号的理论证明,提出了基于盲源提取的背景消减方法,随后基于人工数据和真实数据进行了实验验证,最后提出了基于MSCPE的背景消减方法。

1 视频监视中的盲源提取模型

图1所示为监视视频中的一帧,运动检测的目的是将运动的物体(见图2)从背景(见图3)中

分割。运动的物体也称前景图像,因此,感兴趣的图像(见图1)也可以看作是背景(见图3)和前景(见图2)的线性混合。



图1 源于视频的图片

Fig. 1 An image in a video



图2 从图1中抽取的前景图片

Fig. 2 Extracted foreground image from Fig. 1



图3 从图1中抽取的背景图片

Fig. 3 Extracted background image from Fig. 1

假设二维图像 $f \in \mathbf{R}^{M \times N}$, 通过如下变换将其重构为矢量。

$$\begin{aligned} \mathbf{x}(n) &= \{x[(i-1)N+j] = f(i,j)\} \\ i &= 1, 2, \dots, M; j = 1, 2, \dots, N \\ n &= 1, 2, \dots, M \times N; \mathbf{x} \in \mathbf{R}^{M \times N} \end{aligned} \quad (1)$$

令向量 $\mathbf{s}_f(n)$ 表示前景图像, 向量 $\mathbf{s}_b(n)$ 表示背景图像, 源信号 $\mathbf{s}(n)$ 可以表示为:

$$\mathbf{s}(n) = [\mathbf{s}_f(n), \mathbf{s}_b(n)]^T, \mathbf{s} \in \mathbf{R}^{2 \times MN} \quad (2)$$

感兴趣的视频监视图像 $\mathbf{x}(n)$, 可写为:

$$\begin{cases} \mathbf{x}(n) = \mathbf{A}\mathbf{s}(n) \\ \mathbf{x}(n) = [x_1(n), x_2(n), \dots, x_K(n)]^T \end{cases} \quad (3)$$

式中, $\mathbf{x}(n)$ 表示图像序列 $x_j(n), j = 1, 2, \dots, K; K = M \times N$ 表示一幅图像的像素数; $\mathbf{A} \in \mathbf{R}^{K \times 2}$ 是混合矩阵。

式(3)的物理意义为: 目标图像是背景和前景图像的线性混合。 $\mathbf{x}(n)$ 包含两个图像, 一个是参考图像 $\mathbf{x}_r(n)$, 另一个是感兴趣图像 $\mathbf{x}_i(n)$ 。其线性模型可进一步写为:

$$\begin{cases} \mathbf{x}(n) = \mathbf{A}\mathbf{s}(n) \\ \mathbf{x}(n) = [\mathbf{x}_r(n), \mathbf{x}_i(n)]^T \\ \mathbf{s}(n) = [\mathbf{s}_f(n), \mathbf{s}_b(n)]^T \end{cases} \quad (4)$$

运动检测的目的是通过观察量 $\mathbf{x}(n)$ 提取前景图像 $\mathbf{s}_f(n)$ 。

上述问题可以看作是一个 BSE 模型, 其目标是在模型 $\mathbf{y}(n) = \mathbf{w}^T \mathbf{x}(n)$ 中求解向量 \mathbf{w} 。然而, 源信号之间通常是相关的, 而非独立的。因此, 本文利用相关系数 (Correlation Coefficient, CC) 度量提取信号 s_e 和原始信号 s_o 之间的相关性, 记为:

$$CC = |Cov(s_o, s_e) / \sqrt{Cov(s_o, s_o) Cov(s_e, s_e)}| \quad (5)$$

其中, s_e 和 s_o 具有零均值和标准方差。

表 1 给出了背景图像 (见图 3) 和前景图像 (见图 2) 对应的相关系数, 表明实际图像是相关的。

表 1 背景和前景的 CC 值

Tab. 1 CC values of background and foreground

CC	图 2	图 3
图 2	1.000 0	0.406 6
图 3	0.406 6	1.000 0

传统的 BSS 信号提取方法, 如 PCA 和 ICA 等, 通常在操作前进行去相关, 因此无法一次提取线性混合的相关信号。

2 基于 MSCPE 盲源提取的背景消减法

本节介绍基于 MSCPE^[18,20] 的盲源提取基本原理, 并将其用于前景图像与背景图像相关下的背景消减。

2.1 基于 AR 参数的盲源提取

在盲源提取问题中, N 维随机信号向量 $\mathbf{x}(n)$ 形如 $\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n)$, 其中 \mathbf{A} 为混合矩阵, $\mathbf{s}(n)$ 为源信号向量。盲源提取的目的是在形如 $\mathbf{y}(n) = \mathbf{w}^T \mathbf{x}(n)$ 的式子中求解向量 \mathbf{w} , 从而估计源信号。

为了使算法更加简便快捷, 通常利用预白化处理将观测到的信号 \mathbf{x} 变换为 $\mathbf{H}\mathbf{x}$, 使得 $E\{\mathbf{H}\mathbf{x}\mathbf{x}^T \mathbf{H}^T\} = \mathbf{I}$, 其中 \mathbf{H} 为白化矩阵, $E\{\cdot\}$ 表示求期望。此外, 预白化处理也可以帮助处理病态条件。方便起见, 假设 \mathbf{x} 是预白化处理的结果, 并且具有与 \mathbf{s} 相同的维数。

假设已知期望信号的 AR 参数, p 表示 AR 模型的长度, $\mathbf{e}(n)$ 表示瞬时预测误差 (Prediction Error, PE), 形如:

$$\begin{cases} \mathbf{e}(n) = \mathbf{y}(n) - \mathbf{b}^T \mathbf{y}(n) \\ \mathbf{b} = [b_1, b_2, \dots, b_p]^T \\ \mathbf{y}(n) = [y(n-1), y(n-2), \dots, y(n-p)]^T \\ \mathbf{y}(n) = \mathbf{w}^T \mathbf{x}(n) \end{cases} \quad (6)$$

其中, \mathbf{b} 是期望信号的 AR 参数。

输出 \mathbf{y} 的平均交叉预测误差 (Mean Cross Prediction Error, MCPE)^[22-23] 可以表示为 $E\{\mathbf{e}(n)\mathbf{e}(n-q)\}$, 其中 q 表示误差延迟。相应的 $\mathbf{e}(n)$ 为:

$$\begin{aligned} \mathbf{e}(n) &= \mathbf{y}(n) - \mathbf{b}^T \mathbf{y}(n) = \mathbf{w}^T \mathbf{x}(n) - \\ & [b_1, b_2, \dots, b_p][\mathbf{w}^T \mathbf{x}(n-1), \mathbf{w}^T \mathbf{x}(n-2), \dots, \mathbf{w}^T \mathbf{x}(n-p)]^T \\ &= \mathbf{w}^T \mathbf{x}(n) - \mathbf{w}^T \sum_{i=1}^p b_i \mathbf{x}(n-i) = \mathbf{w}^T [\mathbf{x}(n) - \sum_{i=1}^p b_i \mathbf{x}(n-i)] \end{aligned} \quad (7)$$

假设:

$$\mathbf{z}(n) = [\mathbf{s}(n) - \sum_{i=1}^p b_i \mathbf{s}(n-i)] \quad (8)$$

并且

$$\begin{cases} \mathbf{e}(n) = \mathbf{w}^T \mathbf{A}\mathbf{z}(n) = \mathbf{z}^T(n) \mathbf{A}^T \mathbf{w} \\ \mathbf{e}(n-q) = \mathbf{w}^T \mathbf{A}\mathbf{z}(n-q) = \mathbf{z}^T(n-q) \mathbf{A}^T \mathbf{w} \end{cases} \quad (9)$$

MCPE 可看作 $E\{\mathbf{e}(n)\mathbf{e}(n-q)\}$, 有:

$$\begin{aligned} & E\{\mathbf{e}(n)\mathbf{e}(n-q)\} \\ &= E\{\mathbf{w}^T \mathbf{A}\mathbf{z}(n) \mathbf{z}^T(n-q) \mathbf{A}^T \mathbf{w}\} \\ &= \mathbf{w}^T \mathbf{A} E\{\mathbf{z}(n) \mathbf{z}^T(n-q)\} \mathbf{A}^T \mathbf{w} \end{aligned} \quad (10)$$

假设:

$$\mathbf{Z}(q) = E\{\mathbf{z}(n) \mathbf{z}^T(n-q)\} \quad (11)$$

MCPE 为:

$$E\{\mathbf{e}(n)\mathbf{e}(n-q)\} = \mathbf{w}^T \mathbf{A}\mathbf{Z}(q) \mathbf{Z}^T(q) \mathbf{A}^T \mathbf{w} \quad (12)$$

均方交叉预测误差为 $\mathbf{w}^T \mathbf{A}\mathbf{Z}(q) \mathbf{Z}^T(q) \mathbf{A}^T \mathbf{w}$, 是文献[22]中提取期望信号的代价函数。相应的代价函数的简单形式为:

$$\begin{aligned} & J_q(\mathbf{w}) = \mathbf{w}^T \mathbf{A}\mathbf{Z}(q) \mathbf{Z}^T(q) \mathbf{A}^T \mathbf{w}, \quad 0 < q \leq p \\ & \text{s. t. } \mathbf{w}^T \mathbf{w} = 1 \end{aligned} \quad (13)$$

令 $\mathbf{\Sigma} = \mathbf{Z}(q) \mathbf{Z}^T(q)$, 为对角阵, 其对角元素 $\Sigma(i, i)$ ($i = 1, 2, \dots, M$) 等于源信号 $s(i)$, ($i = 1,$

2, \dots, M) MCPE 的平方。如果源信号的 AR 模型参数不同,则上述代价函数最小为零。因此,可以通过最小化代价函数来提取所需的信号 $\mathbf{J}_q(\mathbf{w})$ 。

由于在约束 $\mathbf{w}^T \mathbf{w} = \mathbf{1}$ 下最小化代价函数 $\mathbf{J}_q(\mathbf{w})$ 等价于求实对称矩阵 $\mathbf{AZ}(q)\mathbf{Z}^T(q)\mathbf{A}^T$ 最小特征值的特征向量, \mathbf{w} 等价于 $\mathbf{Z}(q)$ 最小奇异值的奇异向量。因此有以下算法^[22]:

$$\begin{cases} \mathbf{z}(n) = [\mathbf{s}(n) - \sum_{i=1}^p b_i \mathbf{s}(n-i)] \\ \mathbf{Z}(q) = E\{\mathbf{z}(n)\mathbf{z}^T(n-q)\} \\ \mathbf{w} = \text{MINEVD}\{\mathbf{AZ}(q)\mathbf{Z}^T(q)\mathbf{A}^T\} \\ = \text{MINSVD}\{\mathbf{AZ}(q)\mathbf{A}^T\} \end{cases} \quad (14)$$

其中, $\text{MINEVD}\{\mathbf{M}\}$ 是计算实对称矩阵 \mathbf{M} 的归一化特征向量对应的最小特征值的算子, $\text{MINSVD}\{\mathbf{M}\}$ 是计算矩阵 \mathbf{M} 的归一化奇异值向量对应的最小奇异值的算子。

当存在不同结构时^[25-26], 可以通过最小正规化 MSPE $E\{\mathbf{e}^2(n)\}/E\{\mathbf{y}^2(n)\}$ 来提取源信号。由于 x 是经过白化处理的, 因此去混合标量 $E\{\mathbf{y}^2(n)\}$ 的输出功率是一致的。代价函数可以设为 MSPE, 即 $E\{\mathbf{e}^2(n)\}$, 其表达式如下:

$$E\{\mathbf{e}^2(n)\} = \mathbf{w}^T \mathbf{AR}_p \mathbf{A}^T \mathbf{w} \quad (15)$$

并且

$$\begin{cases} \mathbf{R}_p = E\{[\mathbf{s}(n) - \sum_{k=1}^p b_k \mathbf{s}(n-k)][\mathbf{s}(n) - \sum_{i=1}^p b_i \mathbf{s}(n-i)]^T\} \\ = E\{\mathbf{z}(n)\mathbf{z}^T(n)\} = \mathbf{Z}(q=0) \\ \mathbf{w} = \text{MINEVD}\{\mathbf{AZ}(q=0)\mathbf{A}^T\} \\ = \text{MINSVD}\{\mathbf{AZ}(q=0)\mathbf{A}^T\} \end{cases} \quad (16)$$

其中, \mathbf{R}_p 是对角阵, 并且其对角元素 $R_p(i, i)$ 等于对应源信号的 MSPEE ($e_i^2(n)$)。此外, 当 $q=0$ 时, MSCPE 将退化为 MSPE。

从 BSS 的角度来看, 本质是在被估计的信号 $\mathbf{y}(n) = \mathbf{W}^T \mathbf{x}(n)$ 中求解矩阵 \mathbf{W} 。式(14)对 \mathbf{W} 的求解可以看作

$$\mathbf{W} = \text{SVD}\{\mathbf{AZ}(q)\mathbf{A}^T\} \quad (17)$$

或

$$\mathbf{W} = \text{EVD}\{\mathbf{AZ}(q)\mathbf{Z}^T(q)\mathbf{A}^T\} \quad (18)$$

其中, $\text{EVD}\{\mathbf{M}\}$ 是计算实对称矩阵 \mathbf{M} 的所有归一化特征向量算子, $\text{SVD}\{\mathbf{M}\}$ 是计算实对称矩阵 \mathbf{M} 的所有归一化奇异向量算子。

2.2 相关源的 MSCPE 盲源提取

文献[22]和文献[23]证明, 如果任何期望信号的 AR 参数已知, MSCPE 将会避免排列问题^[22], 并可以提取期望信号^[23]。本文将 MSCPE

推广到以相关源信号为基础, 提取期望信号。

定理 1 在无噪声 BSE 问题中, 观测量 $\mathbf{x}(n)$ 形如 $\mathbf{x}(n) = \mathbf{As}(n)$, 其中 \mathbf{A} 为未知混合矩阵, $\mathbf{s}(n)$ 为源信号向量, 假设:

- 1) 源信号之间是相关的;
- 2) 每个源信号满足:

$$E\{s_i(n)s_i(n-\tau)\} = 0, \tau > 0$$

- 3) 某期望信号的 AR 参数已知;
- 4) 观测量 \mathbf{x} 与源信号具有相同的维度。

那么, 当期望信号的 MSCPE 值小于任意一个源信号时, 式(14)的算法可以有效提取出信号。

证明: 记源信号 $\mathbf{s}(n) = [s_1(n), s_2(n), \dots, s_N(n)]^T$, 假设 $s_1(n)$ 是期望信号, 互相关的源信号在式(19)的限制下, 可以表示为 $\mathbf{s}'(n) = \mathbf{Bs}(n) = [s'_1(n), s'_2(n), \dots, s'_N(n)]^T$ 。

$$\begin{cases} s'_1(n) = s_1(n) \\ E\{s'_i(n)s'_j(n)\} = 0, \forall i \neq j \end{cases} \quad (19)$$

经施密特变换^[13], 有

$$\begin{cases} s_1 = \frac{s_1}{\sqrt{s_1^T s_1}} \\ s_i = s_i - \sum_{j=1}^{i-1} [s_i^T s_j] s_j \\ s_i = \frac{s_i}{\sqrt{s_i^T s_i}} \quad i = 2, 3, \dots, N \end{cases} \quad (20)$$

由假设 2 可知:

$$E\{s'_i(n)s'_j(n-\tau)\} = 0, \forall i \neq j, \tau \geq 0 \quad (21)$$

又由于 $\mathbf{x}(n) = \mathbf{As}(n) = \mathbf{ABs}'(n)$, 为便于后续推导, 假设 $\mathbf{A} = \mathbf{AB}$ 和 $\mathbf{s}(n) = \mathbf{s}'(n)$ 。

当式(18)成立时, $\mathbf{Z}(q)$ 则为对角阵, $\mathbf{Z}(q)\mathbf{Z}^T(q)$ 为非负对角阵, 当 \mathbf{x} 经过白化处理后, 即 $E\{\mathbf{xx}^T\} = \mathbf{I}$, \mathbf{A} 将成为正交矩阵。表示为 $\mathbf{A} = \mathbf{Z}(q)\mathbf{Z}^T(q)$ 。 $\mathbf{AZ}(q)\mathbf{Z}^T(q)\mathbf{A}^T$ 的奇异值分解可表示为: $[\mathbf{A}, \mathbf{A}, \mathbf{A}^T] = \text{SVD}(\mathbf{AZ}(q)\mathbf{Z}^T(q)\mathbf{A}^T)$

对于给定期望源信号 s_k 的 AR 模型参数 \mathbf{b} , 其中 k 为期望数, 各源信号的 MCPE 可表示为^[18] $E\{e_j(n)e_j(n-q)\}$ ($j=1, 2, \dots, N$)。其中 $e_j(n)$ 表示残留误差。由于源信号具有不同的 AR 参数, 有以下特性:

$$E\{e_j(n)e_j(n-q)\} \begin{cases} = 0, j=k \\ \neq 0, j \neq k \end{cases} \quad 0 < q \leq p \quad (22)$$

其中:

$$e_j(n-q) = s_j(n-q) - \sum_{i=1}^p b_i s_j(n-i-q), q \geq 0 \quad (23)$$

其中, p 是 AR 参数的长度, 因此, \mathbf{A} 的第 k 个对角

元素为最小对角元素且为 0。

在约束条件 $w^T w = 1$ 下,最小化代价函数 $J_q(w)$ 等价于求实对称矩阵 $AZ(q)Z^T(q)A$ 的最小特征值对应的特征向量。

因此,如果期望信号的 MSCPE 值小于任意源信号,则式(14)的算法可以提取期望信号。 □

以上定理表明,即使源信号之间互相关, MSCPE 盲源提取方法可以通过最小化 MSCPE 值提取出所需的期望信号。此外,MCPE 盲源提取方法可以通过最小化 MCPE 值提取期望信号。因此,从相关源信号中提取期望信号的关键是,找到合适的参数使相应的 MSCPE 值最小,其中 MCPE 是 MSCPE 的特殊情况。

综上,当期望信号的 AR 参数已知时, MSCPE 盲源提取方法可以从线性混合源中提取期望信号。

2.3 基于盲源提取的背景消减方法

在采用了 MSCPE 的 BSE 算法中,由序列中的参考背景图像和当前场景图像组成数据矩阵,是背景图像和前景图像的混合。

基于 BSE 的背景消减方法包括粗提取和精提取两个阶段。在粗提取阶段,利用 MSCPE 算法将数据矩阵与参考背景图像的 AR 参数进行去混合,得到两幅估计图像:其中一幅是背景图像的精细估计,另一幅是前景图像的粗略估计。在精细提取阶段,利用 MSCPE 算法将数据矩阵与估计的前景图像的 AR 参数进行去混合,估计出精细前景图像。

在粗提取阶段,由于提取的两幅图像互不相关,利用参考背景图像的 AR,背景图像的估计相对较好,前景图像的估计相对粗糙。

如果背景图像和前景图像是独立的,得到的前景图像在粗糙状态下的估计就是解,这两个阶段合并为一个阶段。而背景图像和前景图像通常是相关的,此时两个阶段是必要的。

基于 ICA (BSS) 的方案通常包含两个阶段:训练阶段和检测阶段。与基于 ICA 的方法相比,基于 MSCPE 的方法有如下两个优势:

- 1) 基于 MSCPE 的方法使用的二阶统计量相对于基于 ICA 方法使用的高阶统计量运算开销小;
- 2) MSCPE 无须训练阶段,更加智能。

3 仿真实验

在本节中,使用 MSCPE 提取视频监控中需要的前景图像,以验证该方法的可行性。

3.1 真实图像的人工混合

本实验利用实际数据人工混合来评估

MSCPE 的性能。实验选择了 3 组图片,如图 1~3 所示。图 1 为视频图片,另外两张分别为前景图片和背景图片。然后进行两组实验,分别记为 Exp 1 和 Exp 2。Exp 1 涉及 Image 1 和 Image 2, Exp 2 涉及 Image 1 和 Image 3。

在每个实验中,假设期望信号的 AR 模型参数已知,使用 MSCPE 提取期望信号。另外,背景图像和前景图像将分别作为所需的信号。

实验步骤如下:

Step 1: 选择背景图像和前景图像作为源信号 $s(n)$;

Step 2: 生成混合数据 $x(n) = As(n)$, 其中 $A \in \mathbf{R}^{2 \times 2}$ 表示混合矩阵,由 MATLAB 随机生成;

Step 3: 为提取期望信号选择特定参数 p 和 q , (本文 $p = 60, q = 1$);

Step 4: 利用 MSCPE 提取期望信号,其中期望信号的 AR 模型参数由 MATLAB 中函数“aryule”计算得到;

Step 5: 对提取的信号进行标准归一化,并计算对比相应的 CC。

实验同时使用 MSCPE 和 MSPE 提取期望信号。实验结果如表 2 所示,其中 CC(MSCPE) 是由 MSCPE 求出的提取信号和期望信号计算而得。CC(MSPE) 是用 MSPE 计算的 CC。两种 CC 值都是通过 100 次仿真实验求平均得到。

表 2 Exp1 和 Exp2 的实验结果

Tab.2 Experimental results of Exp1 and Exp2

实验	期望图片	CC(MSCPE)	CC(MSPE)
Exp 1	图像 1	0.999 9	0.205 5
Exp 1	图像 2	0.997 1	0.999 8
Exp 2	图像 1	0.999 4	0.276 6
Exp 2	图像 3	1.000 0	0.988 3

由表 2 可知, MSCPE 提取的图像与期望图像之间的 CC 值均大于 99%, 而 MSPE 在 Exp 1 中图像 2 和 Exp 2 中图像 3 的提取效果较好, 在两个实验中图像 1 的提取效果较差。这是因为 MSPE 来自 ICA, 能够很好地提取独立的源信号, 而表 1 中计算的 CCs 表明, 背景和前景图像不是独立的, 而是相互关联的。因此, ICA 很难从相关源中提取独立的源信号。

对于 MSCPE, 定理 1 的证明表明, 源信号可以变换为 $s'(n) = [s'_1(n), s'_2(n), \dots, s'_N(n)]^T$, 当源信号相互关联时, 只有一个期望

的源信号 $s'_i(n)$ 保持不变。因此, MSCPE 可以逐个提取所需的信号,而传统的 PCA 或 ICA 不能一次提取所有相关的源信号。

备注1 (参数选择): 在 MSCPE 中涉及两个参数 p 和 q 。尽管式(16)表明,在选择不同的 q 时,对于期望信号 MCPE 为零,而 MCPE 的值不为零,甚至在实际中,对于固定的 q ,其值可能并非最小值。这种现象有两方面的原因:一是 AR 参数计算中累积误差的影响;二是源信号是相关的,这使得期望信号的 AR 参数与相关信号的 AR 参数相似。在这种情况下, p 和 q 的选择主要依赖于仿真中的试错。

备注2 (q 的选择): 注意到第 3.1 节中对 q 选择,不同的材料, A 是不同的,而这部分采用 $q = 1$ 。理论上来说, $q > 1$ 不会影响对源信号的估计。但在数值计算中, q 取值越大会要求 BSE 模型的维数较大,进而会增加 SVD 计算的累积误差。因此,需要在 q 和累积误差之间进行权衡,本实验选择 $q = 1$ 。

3.2 单参考背景的真实视频图像

通过第 2 节的理论分析以及第 3.1 节的实验结果表明: MSCPE 可以从相关源中提取期望信号。本节进而利用 MSCPE 在 BSE 框架下,从数据矩阵中提取前景信息。该数据矩阵是由背景图片和当前场景图片组成的序列。

该方法涉及粗提取和细提取两个阶段。

Step 1: 粗提取。采用 MSCPE 算法,利用式(17)或式(18)对参考背景图像的 AR 参数进行去混合。然后用矩阵 $\mathbf{AZ}(q)\mathbf{Z}^T(q)\mathbf{A}$ 最小特征值对应的归一化特征向量将信号分离为背景图像的精估计,另一分离信号为前景图像的粗估计。

Step 2: 细提取。根据第一阶段对前景图像粗估计的 AR 参数,采用 MSCPE 算法,利用式(14)将与第一阶段相同的矩阵进行分层。

在上述两阶段的基础上,分别对单参考背景和多参考背景进行测试。并与图像差分法和粗糙阶段的结果进行对比。

单参考背景即参考背景平稳。在该示例中,仅改变了照明。测试房间有两组荧光灯。灰色的视频图像是 10 帧/s。在第 62 帧之前有一组灯被关闭,之后两组灯都被打开。

图 4 (a) 给出了一个视频序列中的离散场景图像,其帧数分别为 10、85 和 150。选取帧数为 10 的图像作为参考图像,其光照强度与另外两幅检测图像不同。图 4(b)、图 4(c) 和图 4(d) 分别给出了利用两阶段方法、图像差分方法以及粗提取(第一阶段的结果)方法提取到的前景图片。由于帧数为 10 的图像被作为参数,因此对其提取的前景图像的所有像素为 0,即图 4 中第 1 行后 3 幅图像。

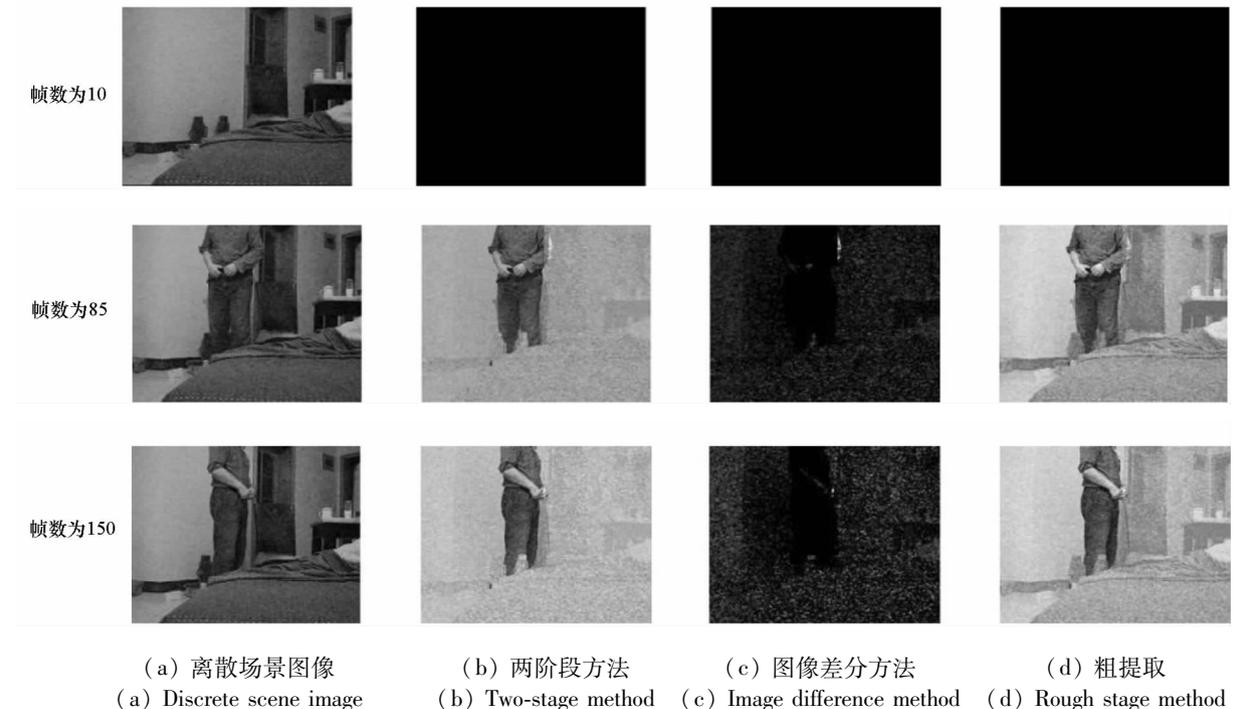


图 4 单参考背景下的实验结果

Fig. 4 Experimental results on single reference backgrounds

图 4(b) 表明,本文提出的两阶段提取方法可以很好地提取期望前景图像。图 4(c) 表明,在参考图像和场景图像光照不同的情况下,传统的图像差分方法效果不佳。图 4(d) 表明,粗估计使得前景图片在背景中相当突出,并且包含了部分背景信息。这是由于背景图像与前景图像相关性的影响。

对图 4 中图像进行二值化,结果如图 5 所示。图 5(b) ~ (d) 分别给出了所提出的两阶段法、图像差分法以及粗糙阶段法的二值化结果。

由图 5 可知,本文所提方法可以在室内光照条件下提取出理想的前景图像。当光线变化时,场景中物体的反射率会发生变化,进而会引起图像大小的改变。值得注意的是,信号大小的变化对信号的 AR 模型参数影响较小。进而说明该方法对光照变化具有较强的鲁棒性。

3.3 多参考背景的真实视频图像

多参考背景是非平稳(动态)的。例如,房间里窗帘的形状经常受到风或行人的影响。第二个室内测试样本是一个人沿着房间的窗帘走。测试房里有两组荧光灯。灰色的视频图像拍摄于 10 帧/s。在第 108 帧之前有一组灯被关闭,之后两组灯都被打开。

图 6(a) 分别给出了第 5 帧、65 帧、145 帧的离散场景图像。帧数为 5 的图像被作为参考图像,因此对其提取的前景图像的所有像素为 0,即

图 6 中第 1 行的后 3 幅图像。图 6(b)、图 6(c) 和图 6(d) 分别给出了利用两阶段方法、图像差分方法以及粗提取(第一阶段的结果)方法提取到的前景图片。图 6(b) 和图 6(d) 表明,本文提出的两阶段提取方法和粗估计方法可以很好地提取期望前景图像。图 6(c) 表明,传统的图像差分方法在多参考背景下效果较差。

对图 6 中图像进行二值化,结果如图 7 所示。图 7(b) ~ (d) 分别给出了所提出的两阶段法、图像差分法以及粗糙阶段法的二值化结果。

由此可知,在多参考背景下,粗糙估计方法和所提出的两阶段估计方法都是可行的。这是由于多重参考背景在式(3)的无噪声模型中引入了噪声。

$$x(n) = As(n) + V(n) \quad (24)$$

其中, $V(n)$ 表示添加的噪声,物理意义为当前背景和参考背景之间的差异。通常有两种处理含噪 BSE 的方法:一是忽略噪声;二是采用噪声消除技术。第二种方法的大部分算法都是基于加性高斯噪声。由于加性噪声难以建模,因此采用无噪声 BSE 算法提取期望信号更为适合。

引入的噪声将不可避免地影响前景图像的 AR 参数估计,进而降低精细估计的效果。然而,本文所提出的方法仍然可以很好地提取期望前景图像。

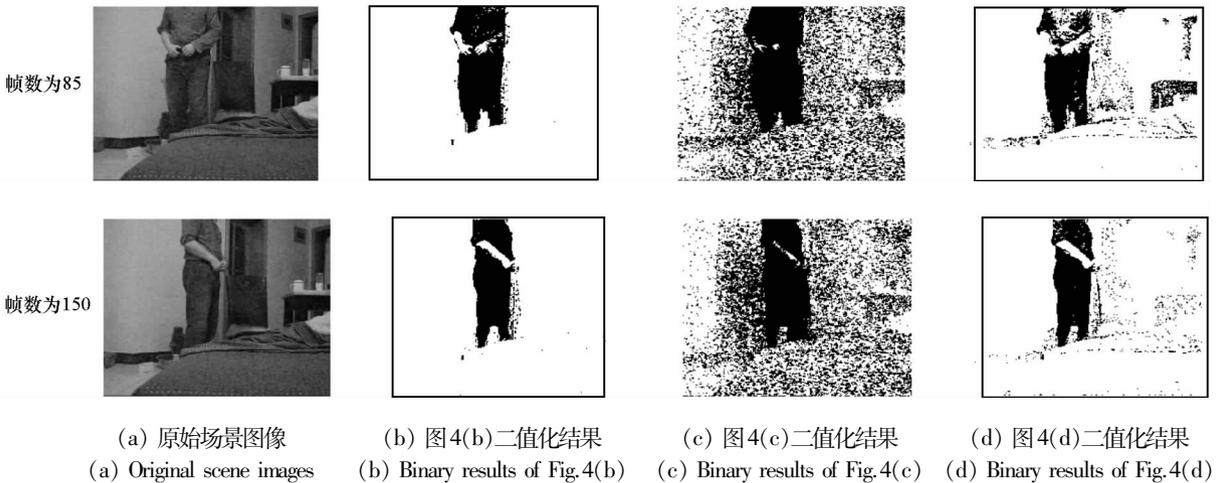


图 5 对图 4 提取的前景图像的阈值化结果

Fig. 5 Thresholding results of the extracted foreground images in Fig. 4

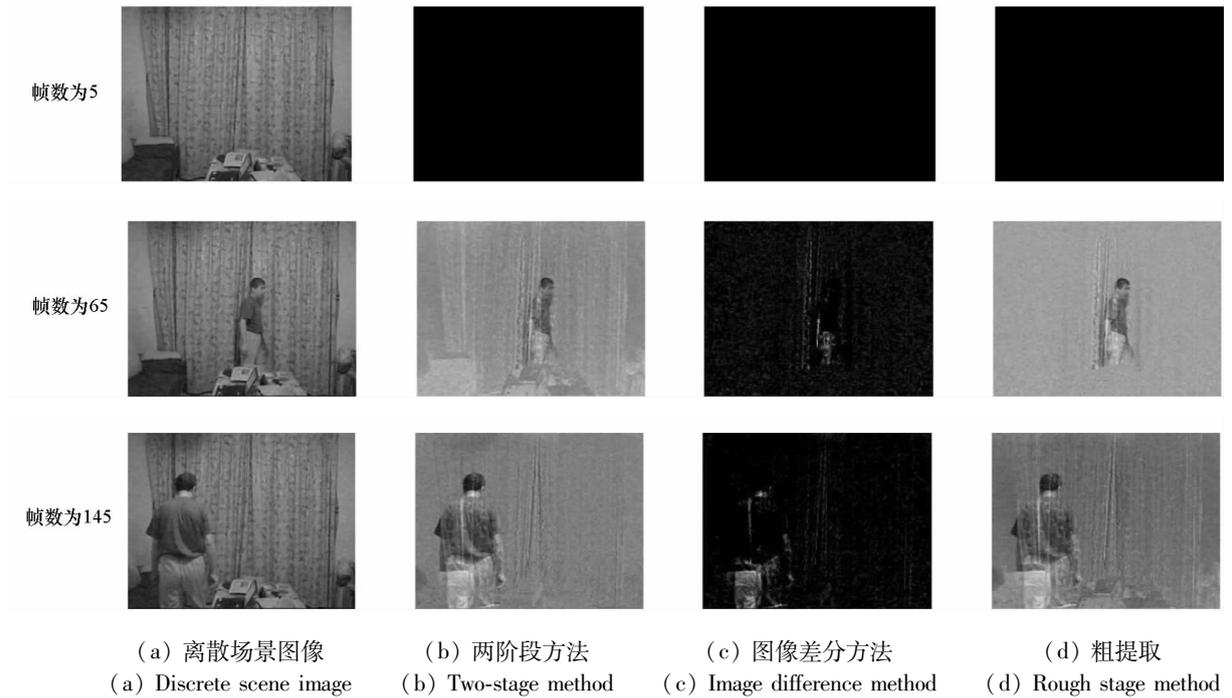


图 6 多参考背景上的实验结果

Fig. 6 Experimental results on multiple reference backgrounds

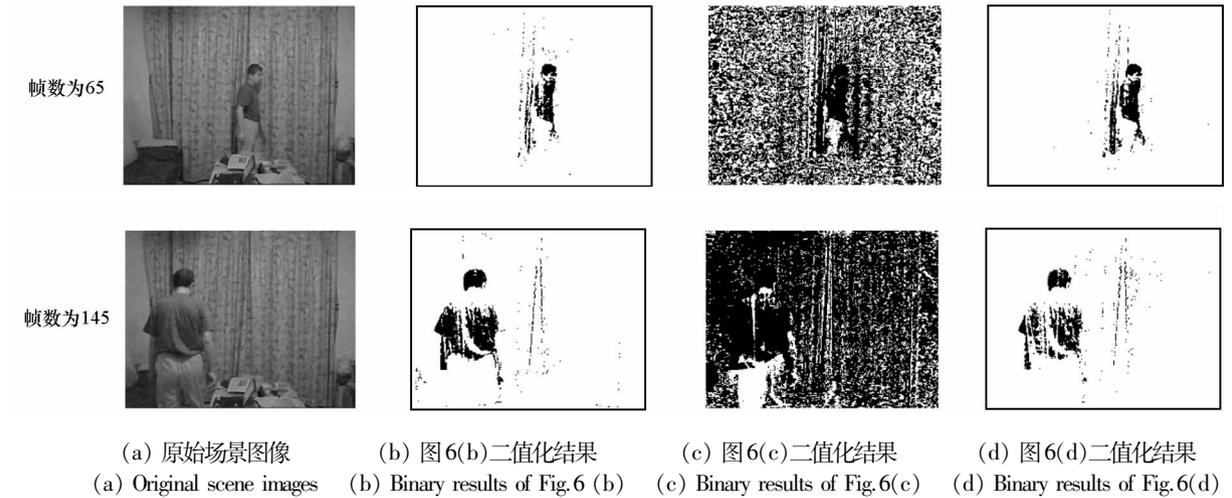


图 7 对图 6 中提取的前景图像的阈值化结果

Fig. 7 Thresholding results of the extracted foreground images in Fig. 6

4 讨论

由于 MSCEP 算法可以从相关源中提取所需的信号,因此该算法也可以用于其他背景消减框架,如基本模型方法和特征背景方法。

4.1 基于 MSCPE 的基本模型

对于基本的背景消减法模型,主要的方法是检测前景对象作为当前帧,与场景静态背景图像之间的差异为:

$$|Frame_i - Background_i| > Threshold \quad (25)$$

解决上述问题的关键是如何自动获取场景的

静态背景图像。一是选择背景作为前 n 帧的平均值或中值^[14],该方法运算速度快,但对内存的开销较大,内存需求是框架大小的 n 倍。二是采用运行平均值:

$$Background_i = \alpha F_i + (1 - \alpha) Background_{i-1} \quad (26)$$

其中: F_i 表示当前场景图片; α 表示学习速率,一般取 0.05。该方法对内存需求不大,但是如果当前场景图片 F_i 包含了前景信息,则性能会严重下降。而 MSCEP 算法可以采用第一阶段粗略提取背景图像,去除或至少减少前景图像的影响,如

图 8 所示。

图 8(a) 和图 8(b) 分别显示背景图像和前景图像的场景图像, 两幅图片被重新排列为数据矩阵 X 。图 8(c) 为在数据矩阵中利用第一阶段方法对前景图片进行粗提取。由图 8(c) 可知, 在保留了背景图片的前提下, 前景图片得到

了弱化。为了进一步验证性能, 对图 8(b) 和图 8(c) 进行零均值标准归一化, 并在图 8(c) 中减去图 8(b) 得到图 8(d), 其二值化结果如图 8(e) 所示。实验结果表明, 当前场景图片 F_t 包含前景图片时, MSCPE 方法仍然能够取得较好的结果。

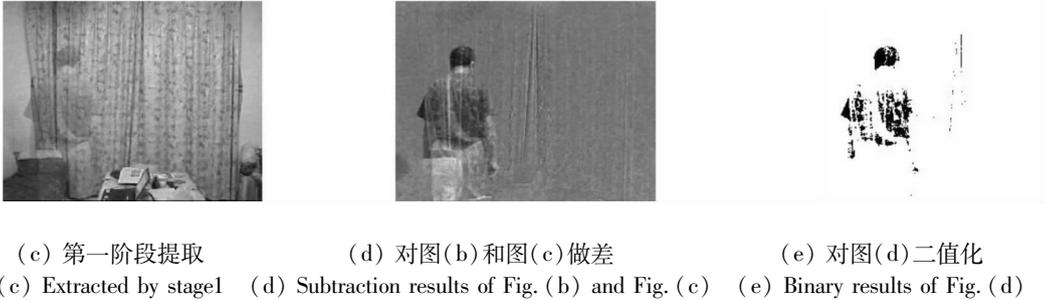
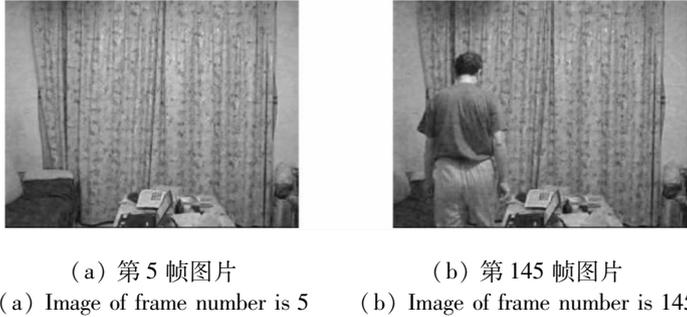


图 8 改进的基本模型对多参考背景的实验结果

Fig. 8 Experimental results of multiple reference backgrounds using modified basic model

4.2 基于 MSCPE 的特征背景模型

对于特征背景模型^[15], 主要是利用主成分分析对 n 帧序列的背景特征进行计算。其主要步骤如下:

- Step 1:** 对 n 帧进行重排列, 得到矩阵 $X \in \mathbf{R}^{n \times l}$, 其中 l 是场景图像的像素个数;
- Step 2:** 计算协方差矩阵 $C = X^T X$, $C \in \mathbf{R}^{l \times l}$;
- Step 3:** 利用协方差矩阵 C , 计算由特征值组成的对角阵 E 以及特征向量所组成的矩阵 Φ ;
- Step 4:** 只保留前 m 个较大特征值对应的特征向量, 将其记为 G , 其中, $G \in \mathbf{R}^{m \times l}$;
- Step 5:** 当前场景图片 $I(1 \times l)$ 可用时, 将其在 G 中进行投影, 并重构为 $I' = I G^T G$;

Step 6: 计算 I 与 I' 的差值, 其差值即为前景图片。

根据特征值的思路, 基于 MSCPE 的特征背景方法的步骤如下:

- Step 1:** 对 n 帧进行重排列, 得到矩阵 $X \in \mathbf{R}^{n \times l}$;
- Step 2:** 利用式(7) ~ (11) 对协方差矩阵进

行计算, $C = AZ(q)A^T$, 其中, $C \in \mathbf{R}^{n \times n}$;

Step 3: 根据式(17) 或式(18), 利用协方差矩阵 C , 计算由特征值组成的对角阵 E 以及特征向量所组成的矩阵 Φ ;

Step 4: 只保留后 m 个较小特征值对应的特征向量, 将其记为 $W(n \times m)$;

Step 5: 记 G 为 m 行特征向量空间, 并由 $G = W^T X$ 计算得到, 其中, $G \in \mathbf{R}^{m \times l}$, 当前场景图片 $I(1 \times l)$ 可用时, 将其在 G 中进行投影, 并重构为 $I' = I G^T G$;

Step 6: 计算 I 与 I' 的差值, 其差值即为前景图片。

图 9 给出了使用修改后特征背景的多个参考背景实验结果, 其中 $m = 4$, $n = 10$ 。图 9(a) ~ (d) 表示最后 m 个特征向量构成一个空间, 图 9(e) 为视频中第 145 帧图片, 对其使用 MSCPE 方法提取的背景图像如图 9(f) 所示, 对图 9(e) 和图 9(f) 做差得到前景图像如图 9(g) 所示, 其二值化结果如图 9(h) 所示。由该图可知, 基于 MSCPE 的特征背景方法可以很好地提取前景图像。仿真实验进一步表明, 基于 MSCPE 的特

征背景方法和原始特征背景方法一样有效。由于篇幅所限,仿真结果未一一列出。

值得注意的是,当 $m = 1$ 时,基于 MSCEP 的特征背景方法将退化为基于 MSCEP 的基本模型方法。因此,基于 MSCEP 的基本模型方法实际上是基于 MSCEP 特征背景方法的一种特例。图 9(c)和图 9(f)表明,基于 MSCEP 的特征背景方法提取的背景图片仍然包含了非前景图片,而基于 MSCEP 的基本模型方法提取的背景图片仍然会受到前景图片的影响。对比表明,扩展子空间对背景图像的提取是有效的。

与此同时,基于 MSCEP 特征背景方法的性能

大部分时候较基于 MSCEP 的基本模型方法好。当 n 帧矩阵 X 包含前景图像时,如果 $m > 1$,特征背景方法就无法工作;当 $m = 1$ 时,基于 MSCEP 的特征背景方法将退化为基于 MSCEP 的基本模型方法,可以正常工作,然而此时特征背景方法效果不佳,如图 10 所示。图 10 分别给出了第 5 帧、145 帧、165 帧构成矩阵 X ,图 10(d)为利用 PCA 得到的重构图片。由图可见,重构图片受到了前景图片干扰。相比而言,MSCEP 能够较好地提取背景图像。然后通过相应的背景消减法 and 二值化结果,验证了基于 MSCEP 的方法优于 PCA 或特征背景法。

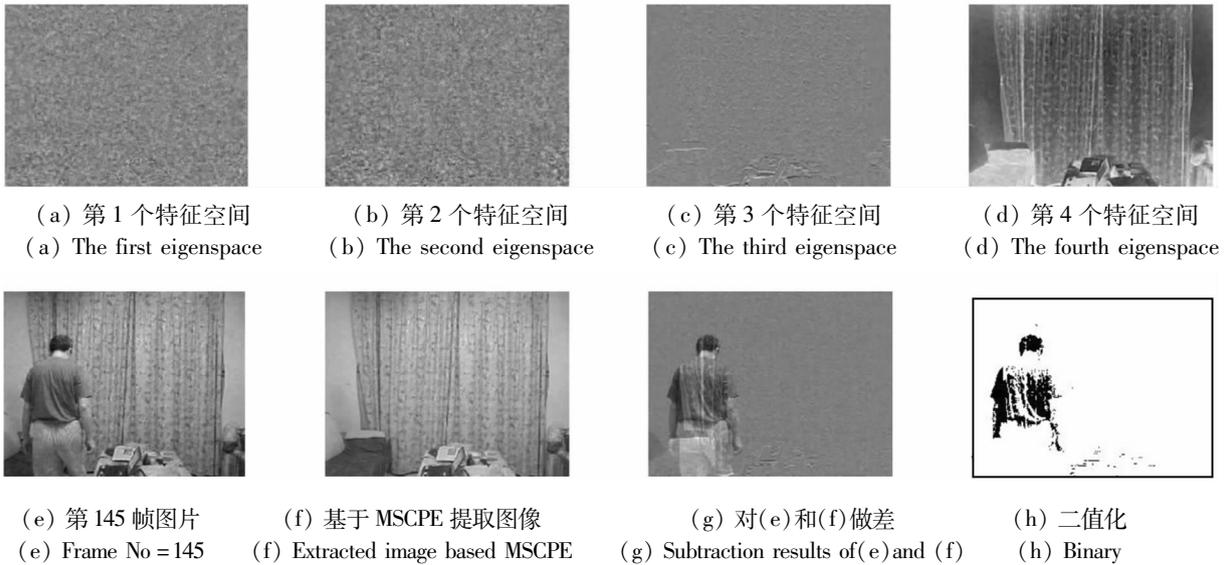
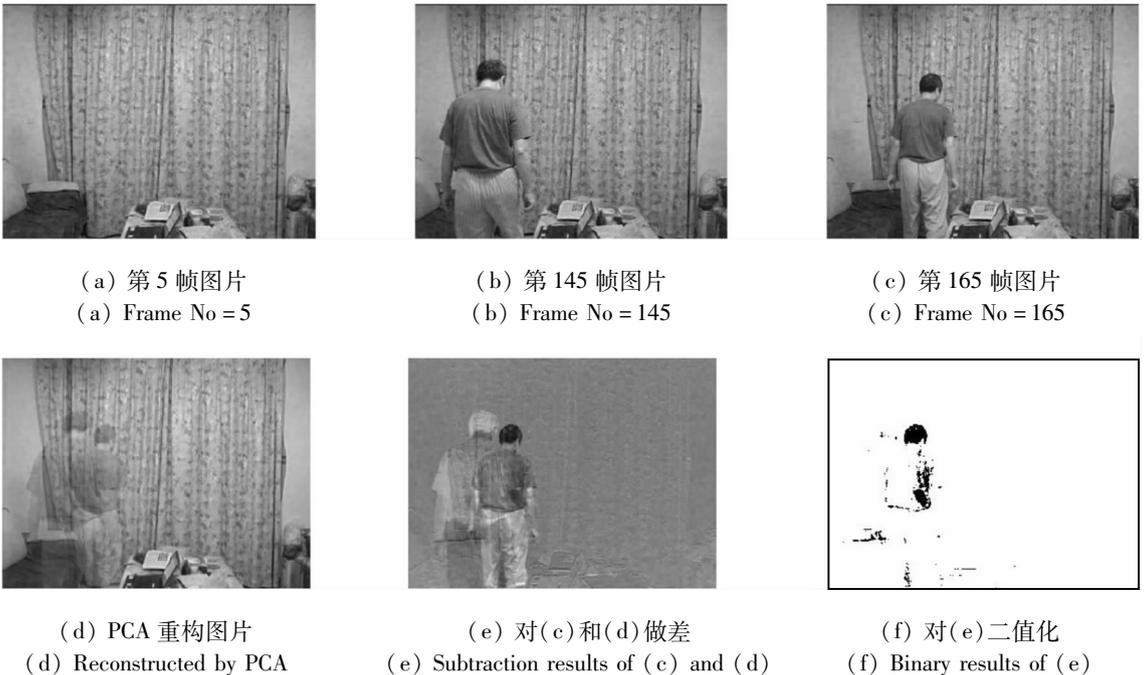


图 9 改进的特征背景对多个参考背景的实验结果

Fig. 9 Experimental results of multiple reference backgrounds using modified eigen-backgrounds



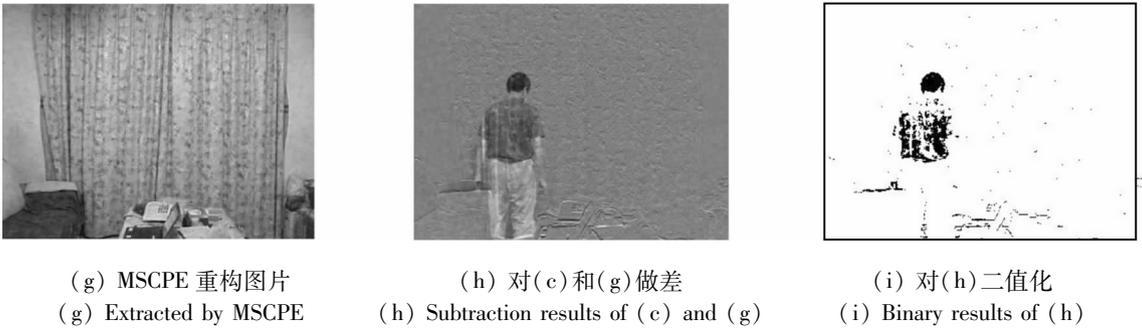


图 10 改进的特征背景对多个参考背景的实验结果

Fig. 10 Experimental results of multiple reference backgrounds using modified Eigen-backgrounds

综上,对于第二类背景消减的改进,有三种方法:基本模型、基于 PCA 的模型和基于 ICA 的模型。MSCPE 算法可以结合这三种方法,提高算法的鲁棒性,进而用于人机交互^[27]、目标跟踪^[28]和人脸识别^[29]问题。

理论上来说,如果对源信号加入适当的约束(非负约束、瞬时结构等^[21]),相关源的准确提取仍然是一个具有很大挑战的问题。除了 BSE 外,其他的线性变换,如邻域线性嵌入^[30]和局部线性嵌入^[31],可能也会被嵌入背景提取中,这将是下一步研究的重点。

5 结论

在视频监控中,一个(帧)场景图像可以被建模为前景和背景视觉内容的线性混合,而实际上图像的背景和前景往往是相互关联的。为了解决从与其他源信号相关的观测数据中提取期望信号的问题,本文将 MSCPE 推广到视频监控的前景图像提取中。通过对实际视频图像的处理,验证了该框架下算法的效果和效能。

在单参考背景下,前景提取可以看作是一个无噪声 BSE 问题。传统的图像差分方法对光照变化很敏感,这是由于图像的大小变化造成的。由于信号的 AR 参数对信号大小的变化不敏感,因此该方法的对光照变化具有鲁棒性。在多参考背景的情况下,后处理模型可以看作是一个有噪声 BSE 问题。由于加性噪声难以建模,因此采用无噪声 MSCPE 算法提取期望信号,在提取期望前景图像时仍表现良好。

MSCPE 还可以用于其他背景消减方案,如基本模型和特征背景模型,从而构成了一个自动化和鲁棒性性能更好的算法框架,可以有效去除或减小前景图像的影响。

参考文献 (References)

- [1] Benzeth Y, Jodoin P M, Emile B, et al. Comparative study of background subtraction algorithms [J]. Journal of Electronic Imaging, 2010, 19(3): 033003.
- [2] Hu W M, Tan T N, Wang L, et al. A survey on visual surveillance of object motion and behaviors [J]. IEEE Transactions on System Man and Cybernetics Part C, 2004, 34(3): 334 - 352.
- [3] Wren C R, Azarbayejani A, Darrell T, et al. Pfinder: real-time tracking of the human body [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 780 - 785.
- [4] Lee D S. Online adaptive Gaussian mixture learning for video applications [J]. Lecture Notes in Computer Science, 2004, 3247: 105 - 116.
- [5] Stauffer C, Grimson W E L. Adaptive background mixture models for real-time tracking [C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1999: 246 - 252.
- [6] Stauffer C, Grimson W E L. Learning patterns of activity using real-time tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 747 - 757.
- [7] Kaewtrakulpong P, Bowden R. An improved adaptive background mixture model for real-time tracking with shadow detection [M]//Remagnino P, Jones G A, Paragios N, et al. Video-Based Surveillance Systems. USA: Springer, 2002: 149 - 158.
- [8] Lee D S. Effective Gaussian mixture learning for video background subtraction [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2005, 27(5): 827 - 832.
- [9] Li L, Huang W, Gu Y H, et al. Statistical modeling of complex backgrounds for foreground object detection [J]. IEEE Transactions on Image Processing, 2004, 13(11): 1459 - 1472.
- [10] Elgammal A, Duraiswami R, Harwood D, et al. Background and foreground modeling using nonparametric kernel density estimation for visual surveillance [J]. Proceedings of IEEE, 2002, 90(7): 1151 - 1163.
- [11] Elgammal A, Duraiswami R, Davis L. Efficient kernel density estimation using the fast Gauss transform with applications to color modeling and tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(11): 1499 - 1504.

- [12] Rai V K, Mohanty A R. Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert-Huang transform [J]. *Mechanical Systems & Signal Processing*, 2007, 21 (6): 2607 – 2615.
- [13] Barnich O, Droogenbroeck M V. ViBe: a universal background subtraction algorithm for video sequences [J]. *IEEE Transactions on Image Processing*, 2011, 20 (6): 1709 – 1724.
- [14] Cucchiara R, Grana C, Piccardi M, et al. Detecting moving objects, ghosts, and shadows in video streams [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, 25(10): 1337 – 1342.
- [15] Oliver N M, Rosario B, Pentland A P. A Bayesian computer vision system for modeling human interactions [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(8): 831 – 843.
- [16] Tsai D M, Lai S C. Independent component analysis-based background subtraction for indoor surveillance [J]. *IEEE Transactions on Image Processing*, 2009, 18(1): 158 – 167.
- [17] Starck J L, Elad M, Donoho D. Image decomposition via the combination of sparse representations and variational approach [J]. *IEEE Transactions on Image Processing*, 2005, 14(10): 1570 – 1582.
- [18] Bobin J, Starck J L, Fadili J M, et al. Sparsity and morphological diversity in blind source separation [J]. *IEEE Transactions on Image Processing*, 2007, 16 (11): 2662 – 2674.
- [19] Herring K T, Mueller A V, Staelin D H. Blind separation of noisy multivariate data using second-order statistics; remote-sensing applications [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2009, 47(10): 3406 – 3415.
- [20] Hyvärinen A, Karhunen J, Oja E. Independent component analysis [M]. USA: John Wiley & Sons, 2001.
- [21] Cichocki A, Amari S. Adaptive blind signal and image processing [M]. USA: John Wiley & Sons, 2003.
- [22] Wang G, Rao N N, Shepherd S J, et al. Extraction of desired signal based on AR model with its application to atrial activity estimation in atrial fibrillation [J]. *EURASIP Journal on Advances in Signal Processing*, 2008, 2008(1): 1 – 9.
- [23] Wang G, Li C G, Dong L. Noise estimation using mean square cross prediction error for speech enhancement [J]. *IEEE Transactions on Circuits & Systems I Regular Papers*, 2010, 57(7): 1489 – 1499.
- [24] Barros A K, Cichocki A. Extraction of specific signals with temporal structure [J]. *Neural Computation*, 2001, 13(9): 1995 – 2003.
- [25] Liu W, Mandic D P, Cichocki A. A class of novel blind source extraction algorithms based on a linear predictor [C]// *Proceedings of IEEE International Symposium on Circuits & Systems*, 2005.
- [26] Liu W, Mandic D P, Cichocki A. Blind second-order source extraction of instantaneous noisy mixtures [J]. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2006, 53(9): 931 – 935.
- [27] Yang Y, Ge S S, Lee T H, et al. Facial expression recognition and tracking for intelligent human-robot interaction [J]. *Intelligent Service Robotics*, 2008, 1(2): 143 – 157.
- [28] Kodagoda K R S, Ge S S, Wijesoma W S, et al. IMM-PDAF approach for road-boundary tracking [J]. *IEEE Transactions on Vehicular Technology*, 2007, 56(2): 478 – 486.
- [29] Zhang B L, Zhang H, Ge S S. Face recognition by applying wavelet subband representation and kernel associative memory [J]. *IEEE Transactions on Neural Networks*, 2004, 15(1): 166 – 177.
- [30] Ge S S, Guan F, Pan Y, et al. Neighborhood linear embedding for intrinsic structure discovery [J]. *Machine Vision & Applications*, 2010, 21(3): 391 – 401.
- [31] Ge S S, Yang Y, Lee T H. Hand gesture recognition and tracking based on distributed locally linear embedding [J]. *Image and Vision Computing*, 2008, 26(12): 1607 – 1620.