一种高精度视频目标检测与分割新方法

赵文哲 秦世引

(北京航空航天大学 自动化科学与电气工程学院, 北京 100191)

摘 要: 根据动态视频场景与多目标检测的应用需求,提出了一种变分光流场与mean shift图像分割相结合的高精度运动目标检测与分割新方法.依据动态场景多运动目标检测的约束条件提出变分光流场优化计算模型,并给出其数值解法.在此基础上提出结合mean shift的高精度运动目标检测与分割算法,此方法对摄像机运动和静止情况都适合,能够进行同一场景中多个运动目标的高精度检测,并且不需要事先的学习和人工干预,具有通用性.

关 键 词: 机器视觉; 运动分析; 目标检测; 运动分割; 变分光流场; 均值平移

中图分类号: TP 391.41

文献标识码: A **文 章 编 号**: 1001-5965(2010)12-1490-05

Novel approach to video object detection and precise segmentation

Zhao Wenzhe Qin Shiyin

(School of Automation Science and Electrical Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Abstract: Taking into account multiple objects detection in dynamic video scene, a method based on variational optimization of optical flow and mean shift was proposed. A variational optimization model for optical flow field was built according to constraints of multiple objects detection in dynamic video scene. The proper numerical solution was then given by this kind of optimizing calculus of variation. With the optical flow optimized solution, a high precise segmentation method was proposed. This approach can be used in the situations of both motion camera and stationary one. Meanwhile it can also be used to detect several targets in dynamic scenes simultaneously without learning in advance and manual intervention so as to be implemented automatically. A lot of experiment results validate the effectiveness of the proposed method.

Key words: machine vision; motion analysis; target detection; motion segmentation; variational optical flow field; mean shift

运动目标检测与分割是计算机视觉研究中的重要内容之一.在视频监控、目标跟踪、运动分析等系统中,往往要求精确地检测和分割出运动目标.常规的运动目标检测方法[1]可以分为:点检测、基于图像分割的检测、背景建模检测和基于监督分类[2]的检测.这些方法处理摄像机静止情况下获取的视频能够取得较好的效果,但对于摄像机运动情况下获取的视频,处理效果就不尽如人意了[3],对于多个运动目标的情况,其检测变得更加困难[4].

文献[5]提出的光流场计算模型,推动了光流场方法的广泛应用,其典型应用在视频压缩领域[6-8].

在最近的十几年中,光流场计算不断涌现新的概念和改进方法.这些改进措施使光流场的计算取得了长足的进步[9],包括提高光流场计算精

收稿日期: 2010-05-10

基金项目: 国家高技术研究发展(863)计划资助项目(2008AA12A200); 国际科技合作资助项目(2007DFA11530); 国家自然科学基金资助项目(60875072)

作者简介: 赵文哲(1979—),女,内蒙古赤峰人,博士生,zwz@asee.buaa.edu.cn.

度的新概念以及改进方法的提出、光流场计算参数对光流场计算结果影响的分析[10]、光流场计算内部机制的洞察,以及对光流场计算精度衡量的量化指标的提出[11]等.光流场计算的变分模型也不断得到改进[12],使用变分法计算光流场在计算精度和运算速度上都显示出超越其他方法的优势[13-14].文献[15]使用运动矢量场与图像分割结合的方法进行动态场景下的目标检测,取得了一定的效果,但分割出的目标轮廓精度不高.

本文针对动态视频高精度检测与分割的应用需求,提出了一种光流场计算的变分模型及其数值解法,首次提出了一种 ODS-VOF/MS(Object Detection and Segmentation based on Variational Optical Flow and Mean Shift)高精度运动目标检测与分割的新方法,解决了动态场景下运动目标精确分割问题.

1　变分光流场计算模型及数值算法

1.1　光流场计算的约束条件分析

光流场是三维运动矢量场在二维平面的投影.为了建立变分光流场的优化计算模型,结合实际应用背景,给出基本假设:①用于计算光流场的前后两帧图像的采样时间间隔足够短;②视频中目标的运动是连续的或分段连续的;③视频拍摄过程中的光照条件是不变的或只有微小变化.下面对这些假设进行模型化分析.

1.1.1　亮度不变性

因为前后两帧图像采样的时间间隔 τ 非常短,拍摄过程中光照条件基本不变情况下,可以认为在前后两帧图像中同一目标所对应的像素点的亮度是不变的,即

$$I(x,y,t)=I(x+u,y+v,t+\tau) \tag{1}$$

其中,$I(x,y,t)$为 t 时刻(x,y)点的亮度值;(u,v)为 $t+\tau$ 时刻(x,y)点的位移增量.

1.1.2　梯度不变性

单一的亮度不变性假设存在一定的缺陷,因为在自然场景中,光照强度会出现微小的变化从而引起亮度值的微小变化,使得基于单一亮度不变性假设的光流场计算模型失效.在亮度出现微小变化时,亮度的梯度能够保持不变[10].因此,假设前后两帧图像中,同一目标所对应的像素点的亮度梯度是不变的,即

$$\nabla I(x,y,t)=\nabla I(x+u,y+v,t+\tau) \tag{2}$$

其中,$\nabla I(x,y,t)$为 t 时刻(x,y)点的亮度梯度值,∇ 为梯度算子.

1.1.3　全局平滑约束

在上述两个条件下,往往使计算得到的光流场存在较多噪声点[2],为了抑制这些噪声点带来的影响,建立全局平滑约束项:

$$E_s=(u,v)=\int_\Omega \psi(|\nabla u|^2+|\nabla v|^2)\mathrm{d}x\mathrm{d}y \tag{3}$$

其中,Ω 为图像的有限支持域;$\psi(t)$为泛函构造函数,其对于总变分最小化过程的收敛有意义.对于图像中的噪声点所在区域,其梯度$\nabla u|^2+\nabla v|^2$往往出现局部值较大的现象,因此在泛函目标函数最小化的过程中,$E_s(u,v)=\int_\Omega \psi(|\nabla u|^2+|\nabla v|^2)\mathrm{d}x\mathrm{d}y$ 也得到最小化,即全局平滑项的引入有效地抑制噪声.但是平滑项在泛函目标函数中占的比重会影响光流场计算的准确性,因此在实际计算中增加正则化系数来控制平滑程度.

1.2　光流场计算的变分模型

根据亮度不变性假设和梯度不变性假设,需要找到最优的$(u(x,y),v(x,y))$,$(x,y)\in\Omega$,使得两帧图像之间的亮度差与亮度梯度差最小化,因此,可以构造出针对图像亮度变差与亮度梯度变差的变分光流场目标函数为

$$E_D(u,v)=\int_\Omega \psi(|I(X+w)-I(X)|^2)+\gamma|\nabla I(X+w)-\nabla I(X)|^2)\mathrm{d}X \tag{4}$$

其中,$X=(x,y)$为图像中的点坐标;$w=(u,v)$为图像中点坐标的位移增量;$I(X+w)$为下一帧图像中像素点 $X+w$ 的亮度值;$I(X)$为当前帧中像素点 X 的亮度值;$\psi(t)$为泛函构造函数;γ 为权重值,其作用是调节能量函数中亮度梯度所占的权重.

引入全局平滑约束项,就可以得到整体优化的泛函目标函数:

$$E(u,v)=E_D(u,v)+\alpha E_s(u,v) \tag{5}$$

其中,α 为正则化系数.

对于 $\psi(t)$的选择,有定理[16]可以证明,构造函数的凸性和在无限远处的线性能保证最小化总变分的问题在有界变差函数空间中的解存在且唯一,凸性对于最小化过程的收敛有意义.本文提出的计算模型中选取 $\psi(t)=\sqrt{t^2+\nu}$,其中 ν 为误差项.显然 $\psi(t)$完全满足变分构造函数的条件.

求解变分光流场优化解的过程就是求解最优的$(u(x,y),v(x,y))$使目标函数 $E(u,v)=E_D(u,v)+\alpha E_s(u,v)$最小.将 Euler-Lagrange 定理应用于本文提出的变分光流场模型,可得

$$\frac{\partial}{\partial u}\psi(|I(X+w)-I(X)|^2+\gamma|\nabla I(X+w)-$$

$$\nabla I(X)|^2)-\mathrm{div}\left\{\frac{\partial}{\partial\nabla u}\psi(|I(X+w)-I(X)|^2+\gamma|\nabla I(X+w)-\nabla I(X)|^2)\right\}+\alpha\frac{\partial}{\partial u}\psi(|\nabla u|^2+|\nabla v|^2)-\alpha\mathrm{div}\left\{\frac{\partial}{\partial\nabla u}\psi(|\nabla u|^2+|\nabla v|^2)\right\}=0 \quad (6)$$

$$\frac{\partial}{\partial v}\psi(|I(X+w)-I(X)|^2+\gamma|\nabla I(X+w)-\nabla I(X)|^2)-\mathrm{div}\left\{\frac{\partial}{\partial\nabla v}\psi(|I(X+w)-I(X)|^2+\gamma|\nabla I(X+w)-\nabla I(X)|^2)\right\}+\alpha\frac{\partial}{\partial v}\psi(|\nabla u|^2+|\nabla v|^2)-\alpha\mathrm{div}\left\{\frac{\partial}{\partial\nabla v}\psi(|\nabla u|^2+|\nabla v|^2)\right\}=0 \quad (7)$$

对式(6)和式(7)进行线性化,可得

$$\psi'(I_z^2+\gamma(I_{xz}^2+I_{yz}^2))(I_zI_x+\gamma(I_{xx}I_{xz}+I_{xy}I_{yz}))-\alpha\mathrm{div}\{\psi'(|\nabla u|^2+|\nabla v|^2)|\nabla u|\}=0 \quad (8)$$

$$\psi'(I_z^2+\gamma(I_{xz}^2+I_{yz}^2))(I_zI_y+\gamma(I_{yy}I_{yz}+I_{xy}I_{xz}))-\alpha\mathrm{div}\{\psi'(|\nabla u|^2+|\nabla v|^2)|\nabla v|\}=0 \quad (9)$$

式中,I_x 为 $I(x,y,t)$ 在 x 方向的一阶偏导数;I_z 为 $I(x,y,t)$ 在 t 上的一阶偏导数;I_{xx} 为 $I(x,y,t)$ 在 x 方向的二阶偏导数;其它符号以此类堆.

对式(8)和式(9)采用迭代求解方法,表达为

$$\psi'((I_z^{k+1})^2+\gamma((I_{xz}^{k+1})^2+(I_{yz}^{k+1})^2))\cdot(I_x^kI_z^{k+1}+\gamma(I_{xx}^kI_{xz}^{k+1}+I_{xy}^kI_{xz}^{k+1}))-\alpha\cdot\mathrm{div}\{\psi'(|\nabla u^{k+1}|^2+|\nabla v^{k+1}|^2)\nabla u^{k+1}\}=0 \quad (10)$$

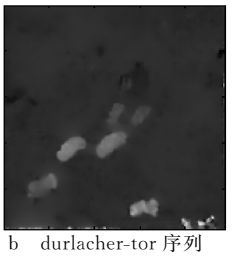
$$\psi'((I_z^{k+1})^2+\gamma((I_{xz}^{k+1})^2+(I_{yz}^{k+1})^2))\cdot(I_y^kI_z^{k+1}+\gamma(I_{yy}^kI_{yz}^{k+1}+I_{xy}^kI_{xz}^{k+1}))-\alpha\cdot\mathrm{div}\{\psi'(|\nabla u^{k+1}|^2+|\nabla v^{k+1}|^2)\nabla v^{k+1}\}=0 \quad (11)$$

式中,上标 k 为迭代次数.

假设一幅图像的尺寸为 $m\times n$,则由每个像素点的迭代方程式(10)、式(11),一共得到 $2\times m\times n$ 个方程组成的大型线性方程组 $\boldsymbol{AQ}=\boldsymbol{b}$,本文采用超松弛迭代法(SOR, Successive Over Relaxation Method)对其进行求解,得到整幅图像的光流场的数值解. 图1所示是序列图像以及其对应的变分光流矢量的模图像. 在光流场模图像中,采用不同的灰度值来表示光流模值的大小. 由光流场模图像可以看出,对于运动速度不同的物体,其光流模值是不同的.



a durlacher-tor 序列



b durlacher-tor 序列光流场模图像

图1 变分光流场计算结果

2 运动目标精确分割

mean shift 的图像分割在本质上是一种通过 mean shift 聚类的无监督图像分割方法. 这种分割方法能够有效地得到图像中物体的轮廓信息.

光流场是包含各像点的瞬时运动速度信息的一个二维向量场. 场景中不同运动速度的物体其光流的值是不同的,根据该特性可知,选取不同的阈值对光流场进行分割,能够得到不同运动速度的物体. 本文提出 ODS-VOF/MS 的运动目标精确分割方法,其基本策略是结合视频帧的运动信息与其轮廓信息,实现运动目标的精确分割. 视频帧图像经过分割后能够得到若干个独立的小区域,利用第1节中得到的视频帧运动信息,确定各个小区域的运动信息,并根据运动速度的不同将有别于背景的小区域分割出来,就能够得到精确的运动目标.

在运行该算法的过程中,需要寻找合适的阈值 T^* 来区分背景与运动目标. 本文利用背景与运动目标的运动信息方差最大化的方法得到阈值 T^*.

综合上述结果,可以给出 ODS-VOF/MS 的高精度目标检测分割算法如下.

输入:短时间隔的两帧视频图像 g_1,g_2.

输出:目标检测与分割的结果图像 g_3.

1) 变分光流计算

① 根据式(10)、式(11)对 g_1,g_2 构造线性方程组 $\boldsymbol{AQ}=\boldsymbol{b}$;

② 使用 SOR 方法得到线性方程组 $\boldsymbol{AQ}=\boldsymbol{b}$ 的解 $\boldsymbol{Q}=[u_{X1},v_{X1},u_{X2},v_{X2},\cdots,u_{Xp},v_{Xp}]^{\mathrm{T}}$;

③ 提取 $\boldsymbol{Q}$ 中每个像素点的光流值($u(x,y)$,$v(x,y)$),并计算其模值 $m_X=\|(u,v)_X\|=\sqrt{u^2+v^2}$,得到光流场模图像.

2) 自适应阈值的确定

① 统计光流场模图像中所有点的亮度级别,记作 n_i;

② 设置初始阈值为 T,将光流场模图像的所

有灰度值分为两组,分别为 $C_0=\{m_{low}\sim T-1\}$ 和 $C_1=\{T\sim m_{high}\}$;

③ 计算 C_0 和 C_1 产生的概率 $\omega_0=\sum_{i=0}^{T-1}p_i=\omega(T)$ 与 $\omega_1=\sum_{i=T}^{m-1}p_i=1-\omega_0$,$C_0$ 和 C_1 的平均值为 $\mu_0=\sum_{i=0}^{T-1}\frac{ip_i}{\omega_0}$ 和 $\mu_1=\sum_{i=T}^{m-1}\frac{ip_i}{\omega_0}$;

④ 计算两组之间的方差 $\delta^2(T)=\omega_0(\mu_0-\mu)^2+\omega_1(\mu_1-\mu)^2$,其中 $\mu=\omega_0\mu_0+\omega_1\mu_1$;

⑤ 从 $\{m_{low}\sim m_{high}\}$ 之间改变 T,求使 $\delta^2(T)$ 达到最大时的 T,记为 T^*.

3) mean shift 图像分割

分别用 x_i 和 z_i,$i=1,2,\cdots,n$ 表示 g_1 和 g_1 分割后图像中的像素点,对每一个像素点:

① 初始化 $j=1$,并且使 $y_{i,1}=x_i$;

② 运用 mean shift 算法计算 $y_{i,j+1}$,直到收敛.记收敛后的值为 $y_{i,c}$;

③ 赋值 $z_i=(x_i^s,y_{i,c}^r)$;

④ 把收敛到同一点的起始点归为一类,把这一类的标号赋给这些起始点,把包含像素点太少类去掉,将每个类之间的边界点赋值 0,其余点赋值 1,就得到分割结果 S_{g_1}.

4) 基于 VOF/MS 的运动目标检测分割

① 对 S_{g_1} 的一个局部区域 i,找到其在光流场模图像的相应区域,统计其中 $m_X>T^*$ 的像素点数占该区域全部像素点数的比率 R_i;

② 若 $R_i>0.5$,则该区域被判定为目标区域;否则该区域被判定为背景区域;

③ 是否所有局部区域都判断完毕,若是,转 4)-④,若否,转 4)-①;

④ 将邻接的背景区域合并;

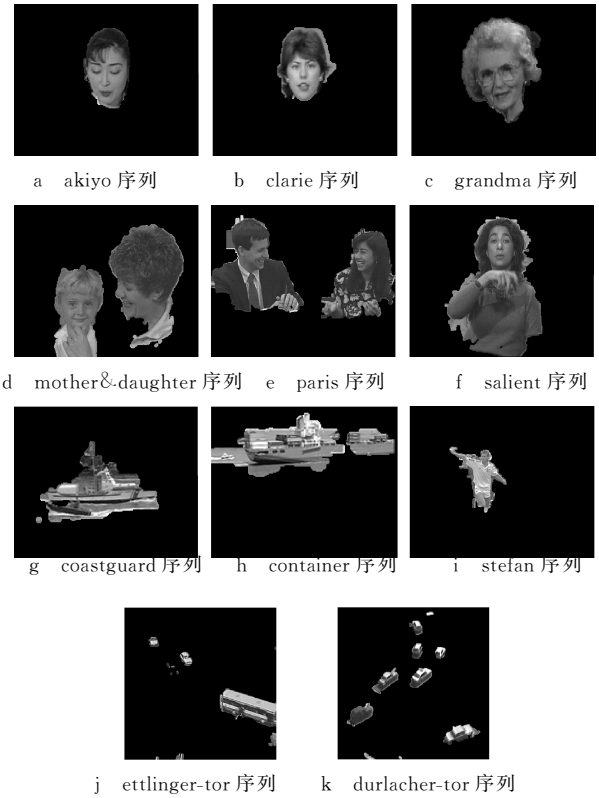
⑤ 将相互邻接的目标区域合并.

该算法在运行过程中,光流场计算的 γ 由视频序列中光线变化程度决定,若视频序列光照变化较大,则选择较大的 γ 值,否则选择较小的 γ 值.SOR 算法的误差门限 ε 的取值影响线性方程组的求解精度,进而影响光流场计算的精度,ε 较小则计算精度较高但耗时长,反之计算精度低但耗时少.α 影响光流场的平滑程度,过大会导致光流场过平滑,过小则光流场包含的噪点较多.

3　实验与结果分析

为了验证本文提出算法的有效性,采用多组视频序列对本文所提出的检测分割算法 ODS-VOF/MS 进行实验验证,采用 MATLAB 在 Pentium(R) 3.0GHz CPU 计算机 WinXP 操作系统上实现该算法,该算法在运行过程中,对于所有试验序列,采用一致的运算参数,分别为:$\gamma=30$,$\alpha=80$,$\varepsilon=10^{-6}$.

图 2a～图 2f 均是摄像机静止情况下拍摄的视频序列,对于此类视频序列,ODS-VOF/MS 算法能够精确地检测并分割出场景中的目标.图 2g、图 2h 为摄像机运动情况下拍摄的视频,但场景中除了包含运动的目标外还包括具有无规则运动的水面以及水波纹,这类对象是目标检测问题的难点所在,因为场景中的无规则运动对象往往使得基于学习的一类方法束手无策,比如背景建模法、帧差分法等.对于这类对象 ODS-VOF/MS 算法分割的轮廓虽然不能精确与目标轮廓完全吻合,如图 2h,但是 ODS-VOF/MS 算法能够检测出运动目标并较好地辨识出目标与无规则运动的水面以及水波纹,从而分割出目标轮廓.图 2i 是摄像机静止情况下拍摄视频,其背景中包含大量的无规则运动(stefan 序列中观看体育比赛的观众的运动),在这种情况下 ODS-VOF/MS 算法能够较好地检测并分割出场景中的目标.图 2j、图 2k 为摄像机静止情况下拍摄的视频,该视频中包含静止的马路上行驶的多部车辆,即多个运动目标,在这种下 ODS-VOF/MS 算法能够精确地检测并分割出场景中的目标.



a　akiyo 序列　　b　clarie 序列　　c　grandma 序列

d　mother&daughter 序列　　e　paris 序列　　f　salient 序列

g　coastguard 序列　　h　container 序列　　i　stefan 序列

j　ettlinger-tor 序列　　k　durlacher-tor 序列

图 2　视频序列的运动目标检测与分割结果

为了进一步测试本文提出算法对摄像机运动时拍摄的视频序列的性能,本文使用了摄像机

运动情况下实拍的视频序列对本文算法做了进一步的测试,其结果如图3所示.可以看出本文算法能够有效检测并分割出运动的目标.



a passerby序列



b bicycle序列

图3 摄像机运动拍摄视频测试结果

为了比较本文提出算法与常规的目标检测方法,本文采用3帧差分法与高斯混合模型法检测了foreman序列中的运动目标,3帧差分法与高斯混合模型检测与分割的目标非常不完整,且由于目标的运动幅度较小,使得大部分目标区域没有检测出来,结果如图4.可以看出,本文算法的检测与分割结果的精度远远高于常规的3帧差分法与高斯混合模型法.



a foreman序列3帧差分法



b foreman序列高斯混合模型法

图4 3帧差分与高斯混合模型检测结果

4 结 论

针对动态视频多目标检测的应用背景,本文根据实际应用的约束条件,建立了光流场计算的优化变分模型,并给出其数值解法,结合mean shift图像分割机制实现了运动目标检测与精确分割.实验结果表明:对于摄像机运动情况与摄像机静止情况下拍摄的视频,ODS-VOF/MS算法都能够有效检测出场景中的单一或多运动目标,且与常规的目标检测方法(如帧差分法、高斯背景建模法等)相比对目标轮廓分割的精度高,在实现动态场景下多运动目标高精度检测与分割方面具有突出的优势.

参考文献(References)

[1] Yilmaz A, Javed O, Shah M. Object tracking: a survey [J]. ACM Computing Survey, 2006, 38(4): Article 13

[2] Stauffer C, Grimson W L. Learning patterns of activity using real-time tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 747-767

[3] 赵文哲,秦世引.视频运动目标检测方法的对比分析[J].科技导报,2009,27(10):64-70
Zhao Wenzhe, Qin Shiyin. Comparative study on detection methods for video motion targets [J]. Science & Technology Review, 2009, 27(10): 64-70 (in Chinese)

[4] Viola P, Jones M, Snow D. Detecting pedestrians using patterns of motion and appearance[J]. International Journal of Computer Vision, 2005, 63(2): 153-161

[5] Horn B K, Schunck B G. Determining optical flow[J]. Artificial Intelligence, 1981, 17: 185-203

[6] Kim S, Han J, Kim J. An efficient scheme for motion estimation using multireference frames in H264/AVC[J]. IEEE Transaction on Multimedia, 2006, 8(3): 457-466

[7] Babu V R, Ramakrishnan K R. Video object segmentation-a compressed domain approach[J]. IEEE Transactions on Circuit and Systems for Video Technology, 2004, 14(4): 462-474

[8] Ewerth R, Schwalb M, Tessmann P, et al. Segmenting moving objects in MPEG videos in the presence of camera motion [C]//Cucchiara Rita. 14th International Conference on Image Analysis and Processing. Washington DC: IEEE Computer Society Press, 2007: 819-824

[9] Brox T, Malik J. Large displacement optical flow: descriptor matching in variational motion estimation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(3): 402-415.

[10] Barron J L, Fleet D J, Beauchemin S S. Performance of optical flowtechniques[J]. International Journal of Computer Vision, 1994, 12(1): 43-77

[11] Galvin B, McCane B, Novins K, et al. Recovering motion fields: An analysis of eight optical flow algorithms[C]// John N Carter. British Machine Vision Conference. London: British Library, 1998: 195-204

[12] Bruhn A. Variational optic flow computation, accurate modelling and efficient numerics[D]. Saarbr ucken: Department of Mathematics and Computer Science, Saarland University, 2006

[13] Brox T, Bruhn A, Papenberg N, et al. High accuracy optical flow estimation based on a theory for warping [J]. International Journal of Computer Vision, 2006, 67(2): 141-158

[14] Sand P, Teller S. Particle video: long-range motion estimation using point trajectories [J]. International Journal of Computer Vision, 2008, 80(1): 72-91

[15] Zhao Wenzhe, Qin Shiyin. A new method for target detection based on analysis of motion vector field in dynamic video scene[C]//Liu Jianguo. Proceedings of the SPIE. Bellingham: SPIE Press, 2009: 74951C-74951C-7

[16] 邹谋炎.反卷积和信号复原[M].北京:国防工业出版社,2001:249-250
Zou Mouyan. Deconvolution and signal recovery[M]. Beijing: National defense industry Press, 2001: 249-250 (in Chinese)

(编 辑:刘登敏)