

融合多特征基于图割实现视频遮挡区域检测

张世辉^{1,2} 何欢¹ 孔令富^{1,2}

¹燕山大学信息科学与工程学院, 河北 秦皇岛 066004

²河北省计算机虚拟技术与系统集成重点实验室, 河北 秦皇岛 066004

摘要 为了准确检测视频中的遮挡区域,提出一种融合多特征基于图割的视频遮挡区域检测方法。基于光流和亮度信息提出三种新的遮挡相关特征—亮度块匹配特征、最大光流差特征和光流残差特征,并定义了所提特征的计算方法。以像素点为单位将所提特征组成特征向量输入随机森林分类器,获取像素点及邻接像素点对的遮挡相关信息。综合利用所获取的遮挡相关信息,通过构造遮挡检测能量函数将遮挡检测问题转化为优化问题。根据该能量函数构造无向图,并基于图割理论对能量函数进行求解,从而得到最终的遮挡区域检测结果。实验结果表明,同现有表现较好的遮挡检测方法相比,所提方法具有较高的准确性和较好的实时性。

关键词 机器视觉; 视频序列; 遮挡区域检测; 光流; 随机森林; 能量函数; 图割

中图分类号 TP391 文献标识码 A

doi: 10.3788/AOS201535.0415001

Fusing Multi-feature for Video Occlusion Region Detection Based on Graph Cut

Zhang Shihui^{1,2} He Huan¹ Kong Lingfu^{1,2}

¹School of Information Science and Engineering, Yanshan University, Qinhuangdao, Hebei 066004, China

²Key Laboratory of Computer Virtual Technology and System Integration of Hebei Province,
Qinhuangdao, Hebei 066004, China

Abstract To detect the occlusion region in video accurately, an occlusion region detection approach is proposed for video by fusing multi-feature based on graph cut. Three new occlusion related features named brightness patch match, maximal flow difference and flow residual are proposed based on the information of optical flow and brightness, meanwhile their calculation methods are defined. The feature vector of each pixel is composed of the proposed features and is inputted into the random forest classifier to obtain the occlusion related information about pixels and adjacent pixel pairs. An occlusion detection energy function, which transforms the occlusion detection problem as an optimization one, is constructed by synthesizing the above occlusion related information. An undirected graph is constructed according to the energy function, then the energy function is solved by graph cut theory to gain the final occlusion region detection result. The experimental results show that, compared with the existing advanced methods, the proposed approach has higher accuracy and better real-time performance.

Key words machine vision; video sequences; occlusion region detection; optical flow; random forest; energy function; graph cut

OCIS codes 150.1135; 150.4620; 110.4153; 100.4145

1 引 言

遮挡既是一种几何现象又是一种光学现象。视频序列中,遮挡是指场景的一部分在当前帧可见,而在下一帧变为不可见的一种现象。遮挡现象的存在会对目标跟踪、运动估计、场景绘制、三维重建、物体识别等视觉任务产生较大影响,故遮挡问题的研究对大部分视觉技术的发展具有重要的基础作用,受到了学者们的广泛关注^[1-4]。

收稿日期: 2014-10-14; 收到修改稿日期: 2014-11-04

基金项目: 国家自然科学基金(61379065)、河北省自然科学基金(F2014203119)

作者简介: 张世辉(1973—),男,博士,教授,主要从事视觉信息处理、模式识别、智能机器人等方面的研究。

E-mail: sshhzz@ysu.edu.cn

视频序列的遮挡检测分为遮挡边界检测和遮挡区域检测两种情况。现有的遮挡区域检测方法主要分为两类:基于模型的方法和基于机器学习的方法。在基于模型的方法中,Strecha等^[5]利用噪声模型和遮挡像素点的亮度直方图信息,提出基于贝叶斯模型的方法来检测遮挡区域。Sun等^[6]对遮挡问题进行建模,在一种离散的情况下进行运动估计和遮挡区域检测。Xiao等^[7]利用水平集的方法将亮度的残差阈值化,从而找到遮挡区域。但上述方法都是通过最小化一个非凸函数来求解问题的,无法保证所求的解为全局最优。Alvarez等^[8]提出一种计算对称光流的遮挡检测方法,即分别计算第一帧图像 I_1 到第二帧图像 I_2 的正向光流,以及第二帧图像 I_2 到第一帧图像 I_1 的反向光流,正反向光流不一致的区域标记为遮挡区域,该方法在处理合成序列时效果较好,但在处理大位移问题上,检测效果并不理想。针对已有方法的不足,Ayvaci等^[9]将传统光流约束方程^[10]的数据项分为可见残差项与遮挡残差项并通过反复加权 L_1 范数将原有问题转化为一个凸问题,从而求得全局最优解,提高了遮挡检测结果的准确性。但该方法的限制在于所建立的凸模型中存在三个调节各项之间权重的正则化系数,这些系数的大小需要人为确定,且没有任何一组系数能保证对所有场景的检测结果都是理想的,使得该方法在通用性上受到了限制。同时,正则化系数确定的问题也存在于大多数基于模型的遮挡区域检测方法中,导致此类方法的通用性较低,进而影响了方法的准确性。

针对基于模型的方法存在的问题,Humayun等^[11]提出了基于机器学习思想的遮挡区域检测方法。该方法首先对图像中各像素点提取遮挡相关特征,然后利用事先训练好的分类器进行遮挡区域检测。Humayun方法分为两个版本:1)采用多特征多维度方法训练分类器的full版本,这个版本的检测结果准确性较高,但由于提取的特征数量较多,维度较大,使得提取特征以及训练分类器的时间都较长,不利于视频中遮挡区域检测的实时性;2)采用特征筛选方法训练分类器的lean版本,此版本由于去掉了特征提取过程中耗时较长的特征,在一定程度上提高了实时性,但降低了准确性。

基于机器学习的遮挡区域检测方法虽然在一定程度上弥补了基于模型方法的不足,但在检测方法的准确性和实时性上仍有待提高。鉴于此,本文综合运用模型和机器学习的方法,采用图割的思想进行遮挡区域检测。提出三种新的遮挡相关特征:亮度块匹配特征、最大光流差特征和光流残差特征;融合这三种新特征,利用随机森林获取遮挡相关信息并构造能量函数;基于图割理论求解该能量函数实现遮挡区域的检测。实验结果表明,与现有表现较好的遮挡区域检测方法相比,所提方法具有较高的准确性和较好的实时性。

2 方法概述

2.1 遮挡区域描述

视频序列中,由于物体之间的相对运动,使得第一帧中的部分区域在第二帧中变为不可见,遮挡区域即为第一帧中变为不可见的那部分区域,如图1所示。其中,图1(a)为视频序列第一帧,图1(b)为视频序列第二帧,图1(a)中箭头所指区域即为本文所要检测的遮挡区域。

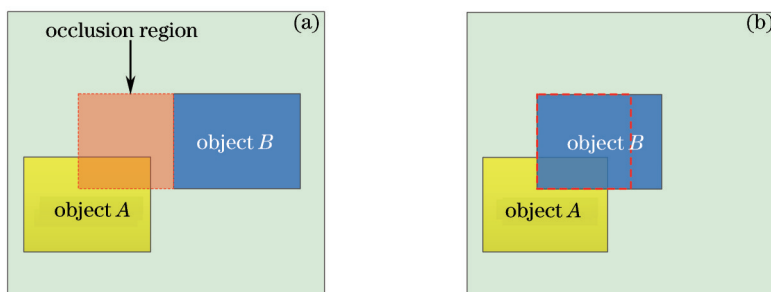


图1 视频序列中遮挡区域示意图。(a)视频序列第一帧;(b)视频序列第二帧

Fig.1 Sketch map of occlusion region in video sequences. (a) First frame in video sequences; (b) second frame in video sequences

2.2 方法总体思想

检测遮挡区域的总体思想如下:1)计算两帧间的光流,并利用两帧间的光流和亮度信息提取第一帧中每个像素点的各遮挡相关特征组成特征向量。2)将各像素点对应的特征向量输入到事先训练好的随机森林分

类器中,得到各像素点的遮挡相关信息,并结合遮挡相关特征进一步获取邻接像素点对的遮挡相关信息。3)综合利用各像素点以及邻接像素点对的遮挡相关信息构造遮挡检测能量函数。4)根据能量函数构造无向图,并应用图割算法求解能量函数得到最终的遮挡区域检测结果。图2为遮挡区域检测方法的总体流程图。

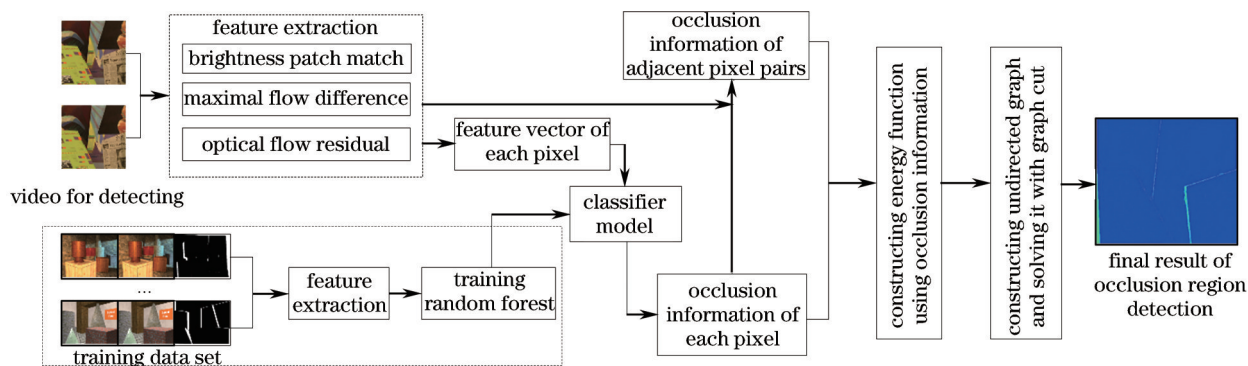


图2 遮挡区域检测方法总体流程

Fig.2 Overall process of the proposed occlusion region detection approach

3 融合多特征基于图割的遮挡区域检测

3.1 新的遮挡相关特征提出及定义

在视频遮挡区域检测过程中,需要对每个像素点进行遮挡相关特征的提取。目前,已有针对视频序列中遮挡区域检测的相关特征主要包括三大类:对应匹配特征、光流属性特征和反向光流特征。其中,对应匹配特征主要有亮度一致性特征和梯度一致性特征。光流属性特征主要有光流梯度特征、光流角变化特征、光流角变化离散度特征、光流长度变化特征、光流长度变化离散度特征以及碰撞速度特征。反向光流特征主要有反向光流一致性特征和反向光流角度一致性特征。但根据已有文献^[11]的实验结果可知,基于上述特征的遮挡区域检测方法在准确性和实时性上仍有待改善,故在对遮挡相关特征深入分析和研究的基础上,提出了属于对应匹配特征的亮度块匹配特征、属于光流特征的最大光流差特征以及属于反向光流特征的光流残差特征,并定义了这三种特征的计算方法。此处需要说明的是,由于光流是特征提取的前提,且存在遮挡、运动边界保持以及大位移等问题,不同的光流算法针对这些问题有不同的侧重点,所以选取了准确性较高且分别应对上述光流问题的三个光流算法文献^[9,12,13]组成光流集 α 作为候选光流集,以多光流为基础进行特征的提取。

3.1.1 亮度块匹配特征

对应匹配特征中的亮度一致性特征和梯度一致性特征均是基于单像素点匹配的特征,当有噪声存在时,这些基于单像素点匹配的特征会因噪声的影响产生较大的误差,从而影响到利用此类特征进行遮挡区域检测结果的准确性。针对这一不足,提出了对噪声更加稳健的亮度块匹配特征,其提取方法为:1)分别对两帧图像以每个像素点为中心构造 3×3 的像素块,对每一个像素块,计算亮度的均值作为该像素点的像素块亮度均值;2)利用光流找到两帧图像间的对应坐标,两坐标对应的像素块亮度均值之差的绝对值即为第一帧图像中该像素点的亮度块匹配特征。亮度块匹配特征 $f_{\text{BPM},\alpha}(\mathbf{x})$ 的定义为

$$f_{\text{BPM},\alpha}(\mathbf{x}) = |P_1(\mathbf{x}) - P_2[\mathbf{x} + \mathbf{u}_\alpha(\mathbf{x})]|, \quad (1)$$

式中 $\mathbf{u}_\alpha(\mathbf{x})$ 是基于光流集 α 中的光流算法计算的光流值,像素点 \mathbf{x} 的坐标为 (i,j) ,则像素点 \mathbf{x} 的像素块亮度均值定义为

$$P_m(\mathbf{x}) = P_m(i,j) = \frac{1}{9} \sum_{x=i-1}^{i+1} \sum_{y=j-1}^{j+1} I_m(x,y), \quad (2)$$

式中 $I_m(x,y)$ 为第 m 帧图像中坐标为 (x,y) 的像素点的亮度。

3.1.2 最大光流差特征

由于大部分遮挡现象发生在光流不连续的地方,此时光流不连续的点就会与其8邻域中至少一个邻域点存在较大程度的光流值阶跃,故提出最大光流差特征用于遮挡区域检测。最大光流差特征是以像素点为单位,通过计算光流图中各像素点与其8邻域像素点光流间的欧式距离,取其中的最大值作为各像素点的最大光流差特征。最大光流差特征 $f_{\text{MFD}}(i,j)$ 的定义为

$$f_{\text{MFD}}(i,j) = \max_{x,y} \|\mathbf{u}(x,y) - \mathbf{u}(i,j)\|, \quad s.t. \quad i-1 \leq x \leq i+1, j-1 \leq y \leq j+1, \quad (3)$$

式中 (i,j) 是光流图中像素点的坐标, $\mathbf{u}(i,j)$ 是坐标为 (i,j) 的像素点对应的光流。

3.1.3 光流残差特征

反向光流特征中的反向光流一致性特征是以坐标的形式进行计算的,通过对坐标进行正反向两次光流的作用即可求得。由于坐标是以整数形式存在的,故将第一帧图像坐标通过光流映射到第二帧图像坐标后,需要对映射后的第二帧图像坐标进行一次取整,这样会产生不必要的误差,而在进行第二帧图像坐标的反向作用后,又需要进行一次坐标的取整,从而使得误差加大。为了避免由于两次坐标取整而产生的计算误差,从光流本身出发,提出了光流残差特征。由于在可见区域内对应坐标下的正反向光流应该正好相反,基于这一思想,计算各像素点从第一帧图像到第二帧图像的正向光流和其对应坐标下第二帧图像到第一帧图像的反向光流,并将两者之和的二范数作为该像素点的光流残差特征。光流残差特征 $f_{\text{FR},\alpha}(\mathbf{x})$ 的定义为

$$f_{\text{FR},\alpha}(\mathbf{x}) = \|\mathbf{u}_{\alpha}(\mathbf{x}) + \mathbf{u}_{\alpha}'[\mathbf{x} + \mathbf{u}_{\alpha}(\mathbf{x})]\|, \quad (4)$$

式中 $\mathbf{u}_{\alpha}(\mathbf{x})$ 和 $\mathbf{u}_{\alpha}'[\mathbf{x} + \mathbf{u}_{\alpha}(\mathbf{x})]$ 分别为坐标 \mathbf{x} 下第一帧图像到第二帧图像的正向光流以及对应坐标下第二帧图像到第一帧图像的反向光流。

3.2 基于所提特征获取像素点及邻接像素点对遮挡相关信息

提取三种遮挡相关特征后,将它们组成特征向量输入随机森林分类器以获取视频帧中的遮挡相关信息。传统的遮挡区域检测方法仅利用单个像素点的遮挡相关信息进行遮挡检测,并没有有效地利用视频帧中像素点与其邻接像素点之间的遮挡相关信息。为了同时利用各像素点及邻接像素点对的遮挡相关信息,首先利用随机森林分类器获取各像素点的遮挡相关信息及初步遮挡标签,然后利用各像素点的初步遮挡标签及遮挡相关特征,进一步获取邻接像素点对的遮挡相关信息。

3.2.1 获取各像素点遮挡相关信息

为了获取各像素点的遮挡相关信息,首先将所提取的各遮挡相关特征以像素点为单位组成特征向量,然后将特征向量输入到事先训练好的随机森林分类器中。由于随机森林分类器是由多棵决策树组成,对于每个像素点,各棵决策树都会对该像素点的遮挡类别进行投票,票数多的类别即为遮挡分类结果,所以通过随机森林分类器,可以得到视频帧中各像素点的决策树投票信息及初步遮挡标签。其中,决策树投票信息即为各像素点的遮挡相关信息。

3.2.2 获取邻接像素点对遮挡相关信息

得到视频帧中各像素点的初步遮挡标签后,为了进一步获取邻接像素点对的遮挡相关信息,对视频帧中每个标签为遮挡的像素点,提取其8邻域内标签为非遮挡的像素点,组成遮挡-非遮挡邻接像素点对。然后计算每一组遮挡-非遮挡邻接像素点对的特征相似性作为邻接像素点对的遮挡相关信息。将遮挡-非遮挡邻接像素点对的特征相似性定义为

$$\varphi(m,n) = |S(m) - S(n)|, \quad (5)$$

式中 (m,n) 为遮挡-非遮挡邻接像素点对, m 为遮挡像素点, n 为非遮挡像素点。 $S(m)$ 和 $S(n)$ 分别为像素点 m 、 n 对应的亮度块匹配特征以及光流残差特征两种特征值之和。

3.3 基于遮挡相关信息构造遮挡检测能量函数并求解

3.3.1 遮挡检测能量函数的构造

为了确定视频帧中各像素点是否为遮挡像素点,综合3.2节中获取的遮挡相关信息构造了一个遮挡检测能量函数。该能量函数反映了将视频帧中一部分像素点划分为遮挡像素点而另一部分像素点划分为非

遮挡像素点时所需要的代价,即将视频帧中所有像素点进行二分类的代价。定义二分类代价包括两个方面:1)将每个像素点判定为遮挡或非遮挡时的代价;2)两个相邻像素点具有不同二分类结果时的代价。所构造的能量函数具体表示为

$$E(\mathbf{y}) = \sum_{i \in R} D_i(y_i) + \sum_{(p,q) \in N} \psi_{p,q}(y_p, y_q) \cdot \delta(y_p, y_q), \quad (6)$$

式中 $\mathbf{y} = (y_1, \dots, y_i, \dots, y_s)$ 是包含 s 个分量的二值向量, s 为待测视频帧中像素点的总数, \mathbf{y} 的每个分量表示相应的像素点是否为遮挡像素点, $y_i = 1$ 时表示像素点 i 是遮挡像素点, $y_i = 0$ 时表示像素点 i 是非遮挡像素点; $D_i(y_i)$ 表示像素点 i 与遮挡或非遮挡像素点的不相似程度,并且将 i 与遮挡像素点的不相似程度作为 i 是遮挡像素点的代价,将 i 与非遮挡像素点的不相似程度作为 i 是非遮挡像素点的代价; $\psi_{p,q}(y_p, y_q)$ 表示将 p, q 两个相邻像素点分为不同类别时需要付出的代价; $\delta(y_p, y_q)$ 是指示函数,当 $y_p \neq y_q$ 时值为 1,当 $y_p = y_q$ 时值为 0; R 是待测视频帧中所有像素点的集合; N 是待测视频帧中所有相邻像素点对的集合。

由于相邻像素点对包含遮挡-遮挡、遮挡-非遮挡、非遮挡-非遮挡以及非遮挡-非遮挡四种情况,根据(6)式的指示函数可知,当 $y_p = y_q$ 时, $\delta(y_p, y_q) = 0$,即相邻像素点对为遮挡-遮挡或非遮挡-非遮挡的情况时, $\psi_{p,q}(y_p, y_q) \cdot \delta(y_p, y_q)$ 的值为 0。而遮挡-非遮挡及非遮挡-非遮挡这两种情况实际为同一组点对中像素点的不同前后顺序,计算代价时只需考虑其中一种情况即可,故(6)式可进一步改写为

$$E(\mathbf{y}) = \sum_{i \in R} D_i(y_i) + \sum_{(m,n) \in N'} \psi_{m,n}(y_m, y_n), \quad (7)$$

式中 (m, n) 为遮挡-非遮挡邻接像素点对, N' 为待测视频帧中所有遮挡-非遮挡邻接像素点对的集合。

$D_i(y_i)$ 具体可利用随机森林中决策树对像素点 i 的投票结果来计算,计算方法为

$$D_i(y_i) = T - C_i(y_i), \quad (8)$$

式中 T 为随机森林中决策树总数, $C_i(y_i)$ 为判定像素点 i 为遮挡类别 y_i 的决策树数量。

$\psi_{m,n}(y_m, y_n)$ 利用遮挡-非遮挡邻接像素点对的特征相似性来衡量将这两个像素点划分为不同类别时需要付出的代价,表示为

$$\psi_{m,n}(y_m, y_n) = \frac{1}{\varphi(m, n)}. \quad (9)$$

由(9)式可知, $\varphi(m, n)$ 的值越小说明相应的两个像素点越相似,这意味着将两个相似的像素点划分为不同类别时需要付出的代价大; $\varphi(m, n)$ 的值越大说明相应的两个像素点不相似,这意味着将两个不相似的像素点划分为不同类别时需要付出的代价小。由于相似的两个像素点应该有相同的分类结果,所以将两个相似的像素点分为不同类别时需要付出更大的代价。此处需要说明的是, $\varphi(m, n)$ 作为分母,其值很小时会导致 $\psi_{m,n}(y_m, y_n)$ 过大,不利于能量函数两项间的平衡,故在对能量函数两项进行研究以及实验分析的基础上,设定一个阈值 $\lambda = 2 \max D_i(y_i)$, 即当 $\psi_{m,n}(y_m, y_n)$ 大于该阈值 λ 时,将 $\psi_{m,n}(y_m, y_n)$ 置为 λ 。

3.3.2 基于图割理论求解遮挡检测能量函数

为了求解遮挡检测能量函数 $E(\mathbf{y})$, 根据需要构造了一个无向图 $G = (R \cup \{s, t\}, E)$, 其中 $R \cup \{s, t\}$ 是顶点集, E 是边集, 如图 3(a) 所示。顶点集中除了包含像素点集 R 外, 还包含两个特殊的顶点 s 和 t , 它们分别代表了遮挡与非遮挡的分类标签。边集 E 中的边也分为两类, 第一类是连接相邻像素点对的边, 第二类是连接像素点与两个特殊顶点 s 和 t 的边。

构造完无向图 G 之后, 根据能量函数 $E(\mathbf{y})$ 对图 G 的各边进行如下赋值: 将像素点顶点 i 与遮挡顶点 s 之间的边 (i, s) 赋权值为 $D_i(0)$, 将像素点顶点 i 与非遮挡顶点 t 之间的边 (i, t) 赋权值为 $D_i(1)$; 将任意两个相邻像素点顶点 p, q 之间的边 (p, q) 赋权值为 $\psi_{p,q}(y_p, y_q)$, 将不相邻像素点顶点之间边的权值赋为 0。

在此基础上, 对遮挡检测能量函数进行求解。首先通过最大流/最小割算法求得图 G 的最小割。图 G 的一个割是指对图中顶点的一个划分, 它把图中所有的顶点划分为两个不相交的顶点集合 S 和 T , 其中遮挡顶

点 $s \in S$, 非遮挡顶点 $t \in T$ 。图 3(b)即为图 G 的一个割。而最小割就是指代价最小的割。在此基础上,若能证明图 G 的最小割对应(7)式中遮挡检测能量函数 $E(\mathbf{y})$ 的最小值,则意味着求解出了遮挡检测能量函数 $E(\mathbf{y})$ 。下面给出图 G 的最小割对应遮挡检测能量函数 $E(\mathbf{y})$ 的最小值的证明。

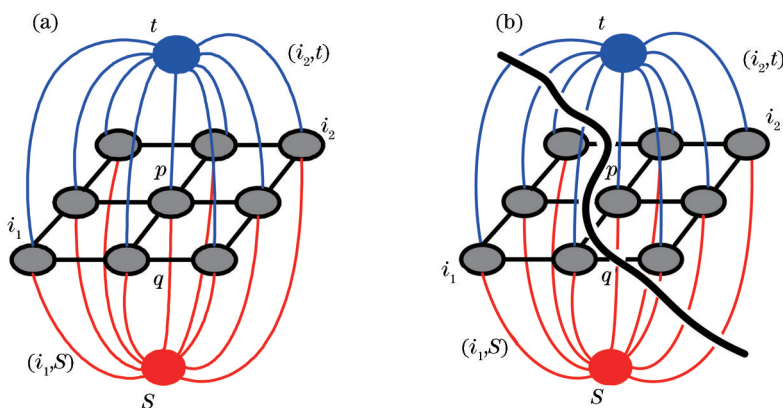


图 3 图 G 及其上的一个割。(a)图 G ; (b)图 G 的一个割

Fig.3 Graph G and a cut of graph G . (a) Graph G ; (b) a cut of graph G

证明:由于图 G 中所有顶点由集合 S 和 T 中所有顶点组成且两者不相交,所以若在像素点集 R 中取两个集合 S' 和 T' ,并令 $S' = \{i | i \in S, i \neq s\}, T' = \{i | i \in T, i \neq t\}$,则有 $S' \cup T' = R, S' \cap T' = \emptyset$ 。由图 3(b)可知,图 G 中割的代价分为三种:1)分割像素点顶点 i 与遮挡顶点 s 之间边 (i, s) 的代价 $c(i, s)$; 2)分割像素点顶点 i 与非遮挡顶点 t 之间边 (i, t) 的代价 $c(i, t)$; 3)分割两个相邻像素点顶点 p, q 之间边 (p, q) 的代价 $c(p, q)$ 。在图 G 的一个割中,当像素点顶点 i 属于某一类别时,则需要将连接该像素点与另一类别顶点的边分割。此外,当两个相邻像素点顶点 p, q 需要分割时,有 $p \in S', q \in T'$ 和 $p \in T', q \in S'$ 两种情况,在无向图 G 中,这两种情况实际为相邻像素点顶点 p, q 的不同前后顺序,分割时只需考虑其中一种情况的代价即可,故图 G 的割的代价可以表示为

$$C = \sum_{i \in S'} c(i, t) + \sum_{i \in T'} c(i, s) + \sum_{p \in S', q \in T'} c(p, q). \quad (10)$$

根据对图 G 中各边所赋的权值可知

$$c(i, s) = D_i(0), \quad (11)$$

$$c(i, t) = D_i(1), \quad (12)$$

$$c(p, q) = \begin{cases} \psi_{p,q}(y_p, y_q), & p \text{ and } q \text{ are adjacent pixel vertex} \\ 0, & p \text{ and } q \text{ are not adjacent pixel vertex} \end{cases} \quad (13)$$

由于 $\sum_{p \in S', q \in T'} c(p, q)$ 中的 $p \in S', q \in T'$,故此处的相邻像素点对 (p, q) 即为遮挡-非遮挡邻接像素点对 (m, n) , 所以

$$\sum_{p \in S', q \in T'} c(p, q) = \sum_{p \in S', q \in T', (p,q) \in N} \psi_{p,q}(y_p, y_q) \cdot 1 + \sum_{p \in S', q \in T', (p,q) \notin N} 0 = \sum_{(m,n) \in N'} \psi_{m,n}(y_m, y_n). \quad (14)$$

又因为能量函数 $E(\mathbf{y})$ 中 $\sum_{i \in R} D_i(y_i)$ 的展开式为

$$\sum_{i \in R} D_i(y_i) = \sum_{i \in S'} D_i(1) + \sum_{i \in T'} D_i(0). \quad (15)$$

将(11)式和(12)式代入(15)式得

$$\sum_{i \in R} D_i(y_i) = \sum_{i \in S'} c(i, t) + \sum_{i \in T'} c(i, s). \quad (16)$$

所以综合(10)式、(14)式和(16)式可知

$$C = \sum_{i \in S'} c(i, t) + \sum_{i \in T'} c(i, s) + \sum_{p \in S', q \in T'} c(p, q) = \sum_{i \in R} D_i(y_i) + \sum_{(m,n) \in N'} \psi_{m,n}(y_m, y_n) = E(\mathbf{y}). \quad (17)$$

将(17)式两端取最小值,即可得图 G 的最小割对应遮挡检测能量函数 $E(\mathbf{y})$ 的最小值,证明完毕。

由于 $E(\mathbf{y})$ 反映了将一部分像素点判断为遮挡而另一部分像素点判断为非遮挡时的代价, $E(\mathbf{y})$ 求得最小值也就意味着此时得到的二分类结果的代价最小,即将每个像素点分为遮挡像素点或非遮挡像素点的结果最准确,从而可最终得到视频帧中遮挡区域的检测结果。

3.4 遮挡区域检测算法描述

算法名称:Occlusion Region Detection Algorithm Based On Graph Cut

Input:视频序列中相邻两帧 I_1, I_2 。

Output: I_1 的遮挡区域检测结果。

Step1:遍历输入的两帧图像 I_1 和 I_2 ,利用光流和亮度信息根据3.1中的方法提取 I_1 中各像素点的亮度块匹配特征、最大光流差特征和光流残差特征。

Step2:将所提取的遮挡相关特征以像素点为单位组成特征向量,输入事先训练好的随机森林分类器中,获得 I_1 中各像素点的遮挡相关信息,并结合所提特征进一步获取邻接像素点对的遮挡相关信息。

Step3:根据所获取的遮挡相关信息,构造遮挡检测能量函数。

Step4:根据遮挡检测能量函数,构造无向图并运用图割算法进行求解,求得 I_1 的遮挡区域。

Step5:可视化 I_1 的遮挡区域检测结果。

4 实验及分析

4.1 实验环境及数据集

为了验证所提方法的检测效果,采用被广泛应用的Middlebury^[14]数据集以及针对遮挡情况较为全面的AlgoSuit^[15]数据集中的视频序列进行实验。实验硬件环境为中央处理器(CPU) Intel(R) Xeon E5-2620,内存16.0G。遮挡区域检测程序采用Matlab软件编程实现。实验过程中,随机森林的参数设定如下:单棵决策树每个节点随机选择的特征数量为4(由于本文采用三种光流算法分别计算所提的三种特征,所以一共得到9个特征,经实验分析得知,从9个特征中随机选择4个特征时效果最佳),决策树中叶子节点大小为20,即当叶子节点中样本数小于20时不再进行分裂,随机森林中决策树的数量为105。

4.2 实验比较及分析

为了全面、合理地评估所提方法的遮挡检测效果,进行了两组对比实验:第一组是所提特征与现有表现较好的特征之间的对比实验;第二组是所提方法与现有遮挡区域检测方法之间的对比实验。

4.2.1 所提特征与已有特征的比较

为了验证所提特征的有效性,应用随机森林分类器对现有遮挡检测过程中表现较好的特征和所提特征在重要性上进行了对比分析。即通过结合现有在遮挡检测中表现好的遮挡相关特征和所提遮挡相关特征同时训练随机森林分类器,训练完成后即可得到各遮挡相关特征的重要性。此为随机森林的优点之一且计算出的各特征重要性大小关系与所选样本无依赖关系,各遮挡相关特征的重要性如图4所示,其中蓝色条纹代表现有特征的重要性,红色条纹代表本文所提特征的重要性。图4中, f_{PC} 为亮度一致性特征, f_{GC} 为梯度一致性特征, f_{BPM} 为亮度块匹配特征, f_{FA} 为光流角变化离散度特征, f_{FN} 为光流长度变化离散度特征, f_{AV} 为光流角变化特征, f_{LV} 为光流长度变化特征, f_{CS} 为碰撞速度特征, f_{TG} 为光流梯度特征, f_{MFD} 为最大光流差特征, f_{RA} 为反向光流角度一致性特征, f_{RC} 为反向光流一致性特征, f_{FR} 为光流残差特征。其中, f_{PC} 、 f_{GC} 和 f_{BPM} 属于对应匹配特征, f_{FA} 、 f_{FN} 、 f_{AV} 、 f_{LV} 、 f_{CS} 、 f_{TG} 和 f_{MFD} 属于光流属性特征, f_{RA} 、 f_{RC} 和 f_{FR} 属于反向光流特征。由图4可知,各遮挡相关特征重要性大小关系为 $f_{FR} > f_{BPM} > f_{PC} > f_{RC} > f_{MFD} > f_{RA} > f_{GC} > f_{TG} > f_{AV} > f_{CS} > f_{LV} > f_{FN} > f_{FA}$,且 f_{BPM} 、 f_{MFD} 与 f_{FR} 都是在同类遮挡相关特征中重要性最高的特征。由此可见,所提出的亮度块匹配特征 f_{BPM} 、最大光流差特征 f_{MFD} 以及光流残差特征 f_{FR} 与现有较好特征相比,所提特征的重要性是比较明显的,从而有利于获得更好的遮挡检测结果。

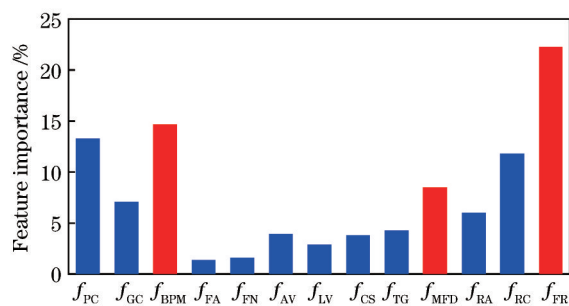


图4 各遮挡相关特征重要性示意图

Fig.4 Sketch map of the importance of all the occlusion-related features

4.2.2 所提方法与现有遮挡区域检测方法的比较

为了进一步评估所提方法的检测效果,将其与现有表现较好的遮挡区域检测方法 Ayvaci^[9]方法和 Humayun^[11]方法进行了比较。其中,Ayvaci方法是基于模型的方法,Humayun方法是基于机器学习的方法。图5给出了三种方法在测试数据集上的部分实验结果,Humayun方法又分为 full 和 lean 版本,因此分别给出了 Humayun方法两版本的检测效果图。图5(a)为原始图像;图5(b)为 Ground Truth;图5(c)为 Ayvaci方法检测结果,标记白色的区域为遮挡区域;图5(d)为 Humayun方法 lean 版本检测结果,标记黄色的区域为遮挡区域;图5(e)为 Humayun方法 full 版本检测结果,标记黄色的区域为遮挡区域;图5(f)为本文方法检测结果,标记绿色的区域为遮挡区域。此处需要说明的是,在视频序列中,部分靠近图像边界的像素点由于光流的作用会移出图像,此时,参考遮挡检测方法^[9,11]对该情况的处理办法,将这部分像素点也标记为遮挡像素点。

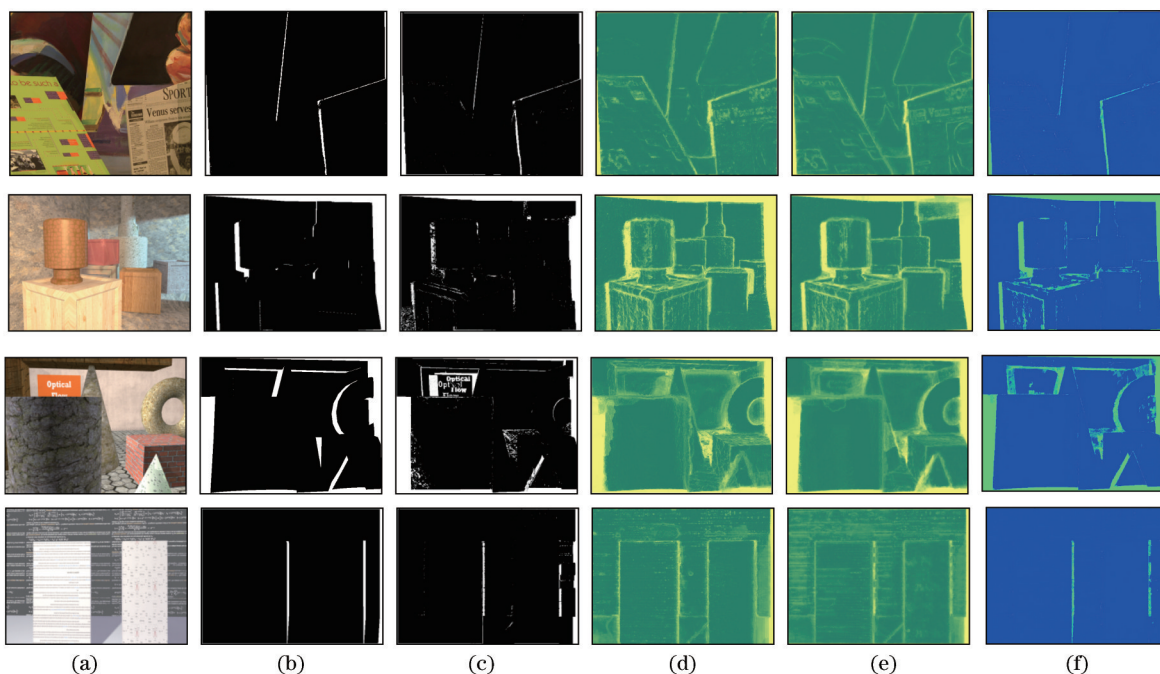


图5 不同遮挡检测方法比较。(a)原始图像;(b)Ground Truth;(c)Ayvaci方法;(d)Humayun(lean);(e)Humayun(full);(f)本文方法
Fig.5 Comparison for different occlusion detection methods. (a)Original image; (b)Ground Truth; (c) Ayvaci method; (d)Humayun(lean); (e) Humayun(full); (f)proposed method

图5第一行为 Middlebury 数据集中 Venus 的检测结果,可以看出,Ayvaci方法可以较为准确地检测遮挡区域,但存在遮挡区域不连续现象;Humayun方法的 lean 版本与 full 版本虽然可以较为全面地检测遮挡区域,但所使用的基于单像素点的亮度一致性特征对噪声比较敏感,因此在两侧报纸中出现了大量的误检区域;本文方法使用了基于区域的像素块亮度匹配特征,增强了检测方法的抗噪能力,保证了检测结果的准确性。第二行为 AlgoSuit 数据集中 Crates1txtr 的检测结果,可以看出,Ayvaci方法虽然可以检测出左侧箱子以

及箱子上圆柱体左侧的遮挡区域,但检测到的遮挡区域并不连续,且未检测出右侧圆柱体附近的遮挡区域; Humayun方法的 lean 版本与 full 版本虽然可以较为全面地检测遮挡区域,但 lean 版本检测到的左侧圆柱体的遮挡区域边界并不清晰,且在图中箱子附近存在明显的误检现象;full 版本的检测效果较 lean 版本有了一定的改善,但由于右上角阴影噪声的影响,使得 full 版本将阴影区域的移动误检为遮挡区域;本文方法使用了抗噪能力较强的亮度块匹配特征及不受亮度匹配影响且准确性较高的光流残差特征,使得检测方法对于阴影的干扰更加稳健,有效地弥补了 Ayvaci方法与 Humayun方法的不足,进一步提高了检测结果的准确性。第三行为 AlgoSuit 数据集中 Brickbox2of 的检测结果,可以看出, Ayvaci方法仅以亮度匹配为准则,因此会将图像左侧文字的移动误检为遮挡区域; Humayun方法的 lean 版本与 full 版本虽未将文字的移动误检为遮挡,但检测出的遮挡区域比较模糊且在边界处存在一些误检区域,特别是右下方圆锥体处;此外, lean 版本在左侧遮挡区域处有明显的误检现象;本文方法使用了可以准确检测遮挡区域边界的最大光流差特征,从而使得遮挡区域的边界不仅清晰而且边界处不存在误检现象。第四行为 AlgoSuit 数据集中 Text1 的检测结果,可以看出, Ayvaci方法在图像右侧纸板附近存在明显的漏检现象; Humayun方法的 lean 版本与 full 版本虽然不存在明显的漏检现象,但由于纸板与背景中存在大量符号,使得基于单像素点匹配特征的 Humayun方法因这些符号噪声的影响产生较大误差,故在背景与纸板中出现了较多误检区域;此外,上述两种方法都将左侧纸板阴影的移动误检为了遮挡区域;本文方法不仅可以有效地检测出遮挡区域,而且可以有效地辨别出阴影的移动,不存在明显的误检现象。因此,与其他两种方法相比,本文方法准确性高、漏检区域少。

图 6 给出了图 5 中三种遮挡检测方法检测结果的查准率查全率(PR)曲线对比图,其中图 6(a)~(d)分别代表图 5 中 Venus、Crates1txtr、Brickbox2of 和 Text1 四组视频序列的结果。由图 6 可知,所提方法的 PR 曲线均优于其他两种方法。

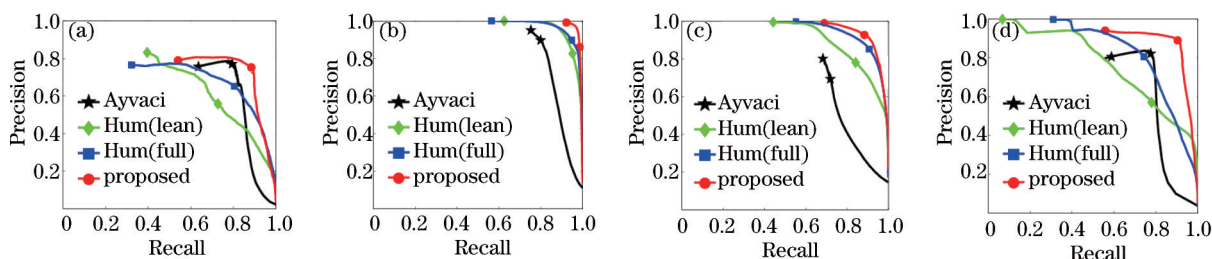


图 6 不同方法的 PR 曲线对比。(a)Venus;(b) Crates1txtr;(c) Brickbox2;(d) Text1

Fig.6 Comparison of PR curve for different occlusion detection methods. (a) Venus; (b) Crates1txtr; (c) Brickbox2; (d) Text1

为了进一步定量评价所提方法的检测效果,对不同方法的查准率 R_p 、查全率 R_r 以及 F-Score [$F-Score = 2R_p R_r / (R_p + R_r)$]进行了比较。表 1 给出了 Ayvaci 方法、Humayun 方法两种版本以及本文方法对 Middlebury 数据集和 AlgoSuit 数据集中全部测试样本进行遮挡区域检测时的平均查准率、平均查全率以及平均 F-Score 的比较结果。由表 1 可知,所提方法在查准率、查全率以及 F-Score 上相比其他方法均有明显提高。

表 1 各遮挡区域检测方法的平均查准率、平均查全率以及平均 F-Score

Table 1 Average precision ratio, average recall ratio and average F-Score for different occlusion region detection methods

Method name	Average precision ratio	Average recall ratio	Average F-Score
Ayvaci	74.30%	64.68%	0.6724
Humayun(lean version)	67.65%	82.25%	0.7192
Humayun(full version)	73.63%	85.29%	0.7684
Proposed method	74.43%	86.22%	0.7795

同时,也将所提方法与同样需要特征提取的 Humayun 方法在特征提取时间、特征提取后的遮挡检测时间及总时间上进行了比较。表 2 给出了 Humayun 方法两种版本以及所提方法对 Middlebury 数据集和 AlgoSuit 数据集中全部测试样本进行遮挡区域检测时的平均特征提取时间、特征提取后的平均遮挡检测时间以及平均总时间的比较结果。由表 2 可知,本文所选特征数量较少且计算简单,因此在特征提取时间、特征提取后的遮挡检测时间以及总时间上均低于 Humayun 方法的 full 版本和 lean 版本,提高了遮挡区域检测的实时性和效率。

表2 两种遮挡区域检测方法的时间消耗比较

Table 2 Time-consumption for two occlusion region detection methods

Method name	Average time-consumption for feature extraction /s	Average time-consumption for occlusion detection /s	Average total time-consumption /s
Humayun(full version)	5997.287	69.224	6066.511
Humayun(lean version)	10.546	54.251	64.797
Proposed method	2.015	46.840	48.855

5 结 论

提出一种融合多特征基于图割的视频遮挡区域检测方法。该方法首先对视频帧中各像素点提取遮挡相关特征,然后将所提特征组成特征向量输入随机森林分类器获取遮挡相关信息并构造能量函数,在此基础上,基于图割理论求解该能量函数,从而实现遮挡区域的检测。主要贡献在于:1)提出三种新的遮挡相关特征—亮度块匹配特征、最大光流差特征和光流残差特征,丰富了视频序列中用于遮挡区域检测的特征种类,且所提特征与现有特征相比重要性较高,使得利用所提特征获取的遮挡相关信息更加准确。2)综合利用了像素点及邻接像素点对的遮挡相关信息,增加了可利用的遮挡信息种类,与传统的仅利用单像素点信息进行遮挡检测的方法相比,提高了遮挡检测的准确性。3)通过构造能量函数将遮挡检测问题转化为优化问题,结合能量函数构造了无向图,并基于图割理论实现了视频中遮挡区域的检测。实验结果表明,同现有表现较好的遮挡检测方法相比,所提方法具有较高的准确性和较好的实时性。

参 考 文 献

- 1 X P Burgos- Artizzu, P Perona, P Dollar. Robust face landmark estimation under occlusion [C]. Computer Vision (ICCV), 2013 IEEE International Conference on. IEEE, 2013: 1513-1520.
- 2 Y Yang, G Sundaramoorthi. Modeling self-occlusions in dynamic shape and appearance tracking [C]. Computer Vision (ICCV), 2013 IEEE International Conference on. IEEE, 2013: 201-208.
- 3 J H Park, H G Lim. Bounding box extraction from spherical hologram of elementary object to synthesize hologram of arbitrary three-dimensional scene with occlusion consideration [J]. Chinese Optics Letters, 2014, 12(6): 060019.
- 4 Wang Wei, Yu Zhijing, Guo Yin, *et al.*. Study on the occlusion problem in dynamic space intersection measurement with multi-camera systems [J]. Acta Optica Sinica, 2014, 34(4): 0415003.
王 伟, 于之靖, 郭 寅, 等. 多摄像机空间交会动态测量中遮挡问题的研究[J]. 光学学报, 2014, 34(4): 0415003.
- 5 C Strecha, R Fransens, L V Gool. A probabilistic approach to large displacement optical flow and occlusion detection [C]. Statistical Methods in video Processing, Springer Berlin Heidelberg, 2004: 71-82.
- 6 J Sun, Y Li, S B Kang, *et al.*. Symmetric stereo matching for occlusion handling [C]. Computer Vision and Pattern Recognition, CVPR 2005, IEEE Computer Society Conference on. IEEE, 2005: 399-406.
- 7 J Xiao, H Cheng, H Sawhney, *et al.*. Bilateral filtering-based optical flow estimation with occlusion detection [C]. Computer Vision-ECCV 2006. Springer Berlin Heidelberg, 2006: 211-224.
- 8 L Alvarez, R Deriche, T Papadopoulos, *et al.*. Symmetrical dense optical flow estimation with occlusions detection [J]. International Journal of Computer Vision, 2007, 75(3): 371-385.
- 9 A Ayvaci, M Raptis, S Soatto. Sparse occlusion detection with optical flow[J]. International Journal of Computer Vision, 2012, 97(3): 322-338.
- 10 B K P Horn, B G Schunck. Determining optical flow[J]. Artificial Intelligence, 1981, 17(1-3): 185-203.
- 11 A Humayun, O M Aodha, G J Brostow. Learning to find occlusion regions [C]. Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011: 2161-2168.
- 12 L Xu, J Jia, Y Matsushita. Motion detail preserving optical flow estimation [J]. Pattern Analysis and Machine Intelligence, IEEE Transaction on, 2012, 34(9): 1744-1757.
- 13 D Sun, S Roth, M J Black. Secrets of optical flow estimation and their principles [C]. Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010: 2432-2439.
- 14 S Baker, D Scharstein, J P Lewis, *et al.*. A database and evaluation methodology for optical flow [J]. International Journal of Computer Vision, 2011, 92(1): 1-31.
- 15 O Mac Aodha, G J Brostow, M Pollefeys. Segmenting video into classes of algorithm-suitability [C]. Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010: 1054-1061.

栏目编辑: 苏 岑