

中图法分类号: TP391.4 文献标志码: A 文章编号: 1006-8961(2011)10-1944-07

论文索引信息: 张欢, 安国成, 张凤军, 王宏安, 戴国忠. 多颜色空间融合的人体检测算法研究 [J]. 中国图象图形学报, 2011, 16(10): 1944-1950

多颜色空间融合的人体检测算法研究

张欢, 安国成, 张凤军, 王宏安, 戴国忠

(中国科学院软件研究所人机交互实验室, 北京 100190)

摘要: 针对运动人体颜色与背景颜色近似时, 单个颜色空间不能很好地刻画人体与背景之间的差别这一问题, 本文提出一种多颜色空间不同颜色分量融合的策略, 即对多个颜色空间的各个分量进行单独的人体检测, 统计人体检测性能最好的若干个通道, 对这些通道的前景进行融合, 得到最终的人体检测结果。另外, 为了处理动态背景问题, 设计一种基于帧间差分法和定时刷新的背景更新算法。实验结果表明, 在复杂场景中, 该算法的检测结果要好于基于单颜色空间的人体检测结果, 并且该算法能够有效地处理动态背景问题。

关键词: 颜色空间; 背景更新; 帧间差分法; 背景减法

Moving human detection algorithm with merging of multiple color space

Zhang Huan, An Guocheng, Zhang Fengjun, Wang Hongan, Dai Guozhong

(Intelligence Engineering Laboratory, Institute of Software Chinese Academy of Sciences, Beijing 100190 China)

Abstract: When the color of a moving human is similar to the color of the background, one single color space can not very well describe the differences between the moving human and the background. Faced with this problem, we propose a multi-color space merging strategy. Moving human detection is conducted separately by finding channels which have the best performance, based on a single channel of color space. Moving human detection then can be concluded by merging these channels. Moreover to solve the dynamic background problem, we designed a background updating algorithm which depends on temporal differencing and refreshing. Experiments confirm that in a complex environment, results of our algorithm are better than test results of the algorithm based on a single color space. The algorithm also can deal with dynamic backgrounds effectively.

Keywords: color space; background updating; temporal differencing; background subtraction

0 引言

近年来,随着计算机视觉的飞速发展,人体检测作为计算机视觉研究的一个热点问题,受到了越来越多的关注,并且被广泛地应用到人机交互、视频监

控、运动分析等多个领域。例如在基于计算机视觉的人机交互中,通常可以采用手势、人体姿态等信息与计算机进行和谐交互,而这类系统往往包括人体检测、人体跟踪以及人体行为识别等模块,其中人体检测是系统的基本构成部分,其性能直接或者间接影响着人体跟踪、人体行为识别的性能,所以人体检

收稿日期: 2011-01-07; 修回日期: 2011-06-27

基金项目: 国家重点基础研究发展计划(973)项目(2009CB320804); 国家高技术研究发展计划(863)项目(2009AA01Z337, 2008AA01Z303); 国家自然科学基金(863)项目(60673188, U0735004); 中国博士后科学基金项目(20100470588); 浙江大学CAD&CG国家重点实验室开放基金项目(A1009)。

第一作者简介: 张欢(1988—), 女。中国科学院软件所人机交互技术与智能信息处理实验室计算机应用技术专业硕士研究生, 研究方向为虚拟现实。E-mail: zhuan44@163.com。

测在基于计算机视觉技术的系统中起到了关键的作用。运动人体检测的实际应用环境的复杂性决定了它对人体运动检测的高可靠性和高实用性的要求,目前该领域的研究还不能完全满足实际应用需求,还需要对其做进一步的研究。

作为计算机视觉研究的重要领域,人体检测得到了国内外学者的高度关注,最近几年,研究人员从不同的角度对其进行了深入研究。Heikkila 等提出利用局部二值模式(LBP)纹理对背景建模^[1],其主要思想是利用像素周围的 LBP 直方图来表示像素,虽然这种背景建模方式具有一定的优点,但是所提取的目标轮廓比较粗糙,具有马赛克效应,而且算法的复杂度较高。Wu 等人提出利用像素时空信息对其进行编码,从而实现动态背景下的前景检测^[2],通过多个动态背景场景,如树木的摇动、水面波动等,进行前景检测均取得了很好的效果。Zha 等人将前景检测看做作为标记问题,提出一种多尺度判别式模式的前景分割算法^[3],首先获得场景的静态背景,然后利用 Boosted 分类器对场景中的周期性运动和真正的运动前景进行划分,由于采用了分类器,所以该算法分割性能受到训练样本的影响,从而使系统的适应能力受到限制。Liu 等人提出一种非参数背景构造方法,用一种有效成分描述的策略对存在移动前景的场景进行背景建模^[4],该算法使用了均值移动算法^[5]进行迭代寻找极值。通过实验可以看出该算法具有抗噪声等优点。但是算法使用了迭代均值移动算法,其很难保证迭代达到全局最优。Cheng 等人提出在线估计基于混合高斯的背景模型的高斯成分个数的方法^[6],该方法具有高效以及场景适应能力强等特点。但也不能适用光照突然变化等情况。Chen 等人提出将基于像素级别和基于块的背景建模方法统一的框架^[7],从而达到两类背景建模方法的互补,但是系统如何有效地将这两个层面的处理结构进行有效的融合还需要进一步研究。Wang 等人提出利用背景像素的颜色信息和运动信息对前景进行检测^[8],但是该系统存在明显的缺陷,即假设被检测目标是直立的行人、前景在遮挡时外观要发生明显变化等,这些前提假设在很多实际应用中是不满足的,所以系统的进一步实用化还需要进行研究。Zhang 等人针对前景颜色与背景颜色类似的问题,提出利用前景历史运动信息对前景进行有效分割^[9],但是当前景突然改变运动模式,该算法性能将明显下降,所以其实用性也是具有一

定的局限性。Li 等人通过支持向量机完成了快速准确的人体检测,其主要思想是结合级联拒绝的机制与梯度方向直方图特征^[10]。Xu 等人结合明暗度、颜色及边缘信息,利用隐马尔可夫解码算法完成人体检测^[11]。Song 等人提出利用像素的饱和度(S)、亮度(V)值的变化,从而完背景的更新和重建过程^[12]。

在人体检测过程中,伪装问题,即运动人体颜色与背景颜色相似导致难以正常检测出运动人体的问题,通常是比较常见的,这是因为基于计算机视觉系统的人体检测主要是检测人体衣服的颜色,而人体衣服的颜色很有可能与背景的颜色相似,如身穿绿色衣服坐在草地上的人、身穿蓝色衣服在海边行走的人。人体检测的另一个难点问题是在动态背景,即背景在一段短暂时间中处于运动状态。和伪装一样,动态背景在生活中也是很常见的,如打开和关闭的门,拿走桌子上的一本书等。人体检测的理想情况是只检测出人体,其余的均以背景形式存在,而伪装问题和动态背景问题的存在会导致无关物体被检测出,甚至可能造成运动人体的误检。为解决伪装和动态背景问题而提出多颜色空间融合算法,该算法包含两个模块:一个模块解决伪装问题;另一个模块解决动态背景问题。在存在伪装问题的情况下,基于单个颜色空间的人体检测结果不是很理想。本文算法大致步骤是通过找到多个颜色空间的多个颜色通道中检测运动人体效果最好的若干通道,对这若干通道的前景进行融合,得到完整的人体检测结果。而在存在动态背景问题的情况下,本文提出的背景更新算法能够及时地将动态背景更新到背景中去。

1 基于多颜色空间融合的人体检测算法

常用的人体检测算法有 3 种,分别为背景减除法、帧间差分法和光流法。背景减除法是目目前最常用的一种方法,能够快速、有效、完整地检测出目标图像,缺点是对运动场景的适应性不好。帧间差分法的优点是对运动场景的适应性较好,缺点是检测出来的目标不完整。光流法的优点是在摄像头运动的情况下也能检测出运动目标,缺点是计算复杂、抗噪性能差。考虑到本文研究的视频是处在摄像头固定和室内光照变化缓慢的环境下,选用背景减除法作为人体检测算法。

从视频中得到一帧一帧的图像。为了消除噪声,对图像进行预处理,考虑到高斯类型噪声的存在,选用高斯滤波器作为预处理方式。基于帧间差分法进行背景建模。背景建模的具体方法是,如果某一像素被帧间差分法检测为前景,则不进行背景更新,否则将该像素按照一定的频率更新到背景中去。完成背景建模后,基于背景减法进行人体检测,背景也同时进行更新。由于检测结果中仍然存在噪声,则需要对检测结果进行形态学的腐蚀处理。又因为光照不均衡等会造成检测出来的人体轮廓不完整,则还需要对腐蚀后的检测结果进行形态学膨胀处理。对不同颜色空间的不同颜色通道的检测结果进行融合,得到最终的人体检测结果,具体流程如图 1 所示。

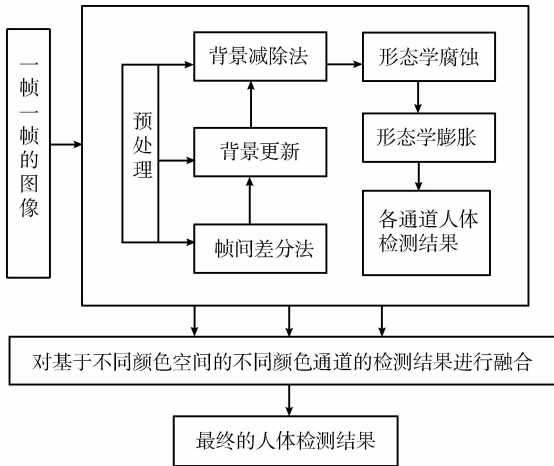


图 1 人体检测流程图

Fig. 1 Human detection flowchart

1.1 不同颜色空间融合

颜色空间通常是由 3 个颜色通道组成的,不同的颜色空间对颜色有不同的刻画能力,并且不同的颜色通道反映了不同颜色空间的不同特性,并不存在一个能够完全刻画颜色的空间。本文提出的多颜色空间融合的算法,即选出检测性能最好的 3 个颜色通道,组成颜色通道向量,将它们的检测前景进行融合,得到最终的检测结果。选择的颜色空间为 RGB、HSV 和 YCrCb。

RGB 作为最常用的颜色空间,也是一般的图像和视频的存储方式。RGB 由 3 个通道 R(红色)、G(绿色)和 B(蓝色)组成,其中 R、G、B 的取值范围均为 0 ~ 255。HSV 作为面向视觉感知的颜色空间,由 H(色调)、S(饱和度)和 V(亮度)组成,其中

H 的取值范围为 0 ~ 360°,S 和 V 的取值范围均为 0 ~ 1。YCrCb 即 YUV,Y 表示明亮度,Cr 表示 RGB 中不计亮度值的红色分量,Cb 表示 RGB 中不计亮度值的蓝色分量,其中 Y、Cr、Cb 的取值范围均为 0 ~ 255。YCrCb 的特点是它的亮度信号 Y 和色度信号 Cr、Cb 是分离的。

在对视频的初始处理中,从视频中得到了一帧一帧的图像,此时的图像是 RGB 形式,将其分解为 R、G、B 这 3 个通道,按照颜色空间的转换公式,得到 R、G、B、H、S、V、Y、Cr、Cb 这 9 个颜色通道的值。颜色空间之间的转换如下:

1) RGB 转换为 HSV

$$\max = \max(R, G, B)$$

$$\min = \min(R, G, B)$$

$$\text{if } R = \max, H = (G - B) / (\max - \min)$$

$$\text{if } G = \max, H = 2 + (B - R) / (\max - \min)$$

$$\text{if } B = \max, H = 4 + (R - G) / (\max - \min)$$

$$H = H \times 60$$

$$\text{if } H < 0, H = H + 360$$

$$V = \max(R, G, B) \tag{1}$$

$$S = (\max - \min) / \max \tag{2}$$

2) RGB 转换为 YCrCb

$$\begin{cases} Y = 0.299R + 0.587G + 0.114B \\ Cr = 0.5R - 0.4187G - 0.0813B + 128 \\ Cb = -0.1687R - 0.3313G + 0.5B + 128 \end{cases} \tag{3}$$

对 R、G、B、H、S、V、Y、Cr、Cb 这 9 个通道分别进行人体(图 2)检测,结果如图 3 所示。

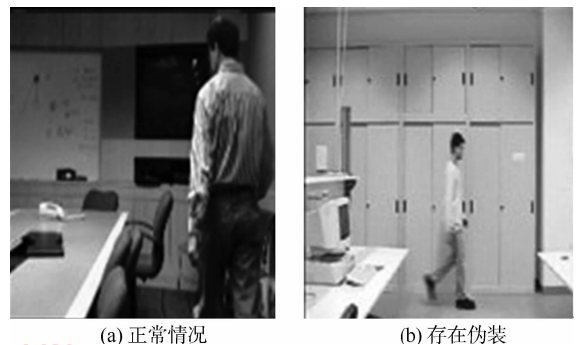


图 2 视频中的一帧

Fig. 2 A frame in the video

图 3(a)为图 2(a)的人体检测结果,图 3(b)为图 2(b)的人体检测结果。图 2(a)中运动前景的颜色与背景的颜色差异较大,则运动人体比较容易被检测出来,而图 2(b)中运动前景的衣服颜色与背景

颜色相似,即存在伪装问题,运动人体不容易被完整地检测出来。无论是图2(a)还是图2(b),R、G、B、Y通道都能够较好地完成人体检测,但是这4个通道中不能被检测出来的人体部分也基本相同,在图3(a)中表现为对运动人体背部的检测结果不完整,在图3(b)中表现为不能检测出运动人体的上身部分,且对人体腿部的检测效果不够好,检测结果中会出现两条腿重叠现象。H通道在图3(a)不能检测出运动人体的头部,在图3(b)中只能检测出腿部,检测效果都比较差。Cr通道在图3(a)中效果比较好,但是在图3(b)中效果却很差。对于S、V、Cb通

道,在图3(a)中,V和Cb能够得到比较完整的头像,S、Cb能够得到比较完整的身子,S、Cb能够得到比较完整的腿部,在图3(b)中,S和V能够得到比较完整的头像,S、Cb能够得到比较完整的身子,V、Cb能够得到比较完整的腿部。综上所述,认为选取颜色通道S、V、Cb构成通道向量,对它们的检测结果进行融合,即使存在伪装问题,也能够得到比较完整的人体检测结果。融合的方法是将某像素至少在两个通道中被检测为前景,则为前景,否则为背景。经过融合操作的检测结果能够有效地消除S通道的噪声。

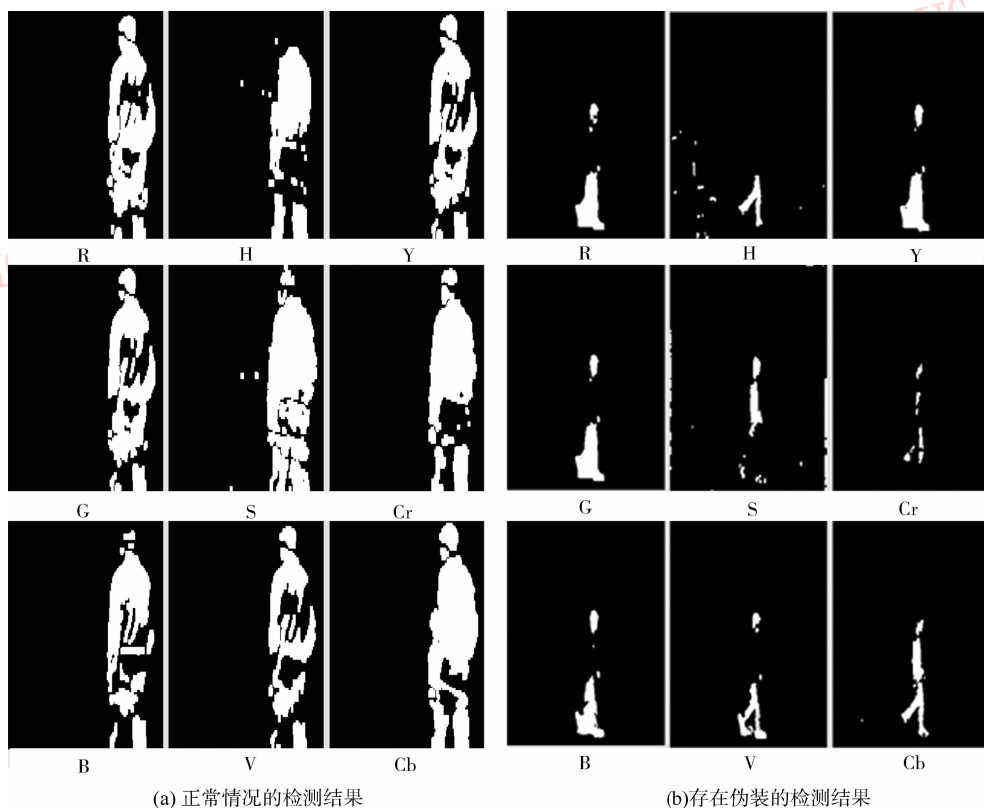


图3 9个通道的人体检测结果

Fig. 3 Test results of 9 channels

1.2 背景更新

为了抑制光照变化以及动态背景问题对人体结果造成的影响,系统需要进行背景更新。传统的背景更新方法是将所有像素按照一定比率更新到背景中,这个更新率一般较小,因为更新率太大会造成人体检测结果不好。这就导致了传统更新算法将场景中的静止物体更新到背景中的速度较慢。在存在动态背景的情况下,如果系统不能及时将已经处于静止状态的动态背景更新到背景中,则动态背景会在

很长一段时间内被当做前景给检测出来。当人体经过该动态背景的时候,会出现因为动态背景和运动人体的检测结果重合而不能正确检测出人体的情况,如图4、图5所示,其中图5分别为图4对应的基于一般更新算法的人体检测结果。

帧间差分法的特点是能够有效地检测出运动物体而检测不出静止的物体。动态背景在经过短暂的运动后静止,由于之前背景的更新,动态背景能够被背景减法检测出来,但是不能被帧间差分法检测

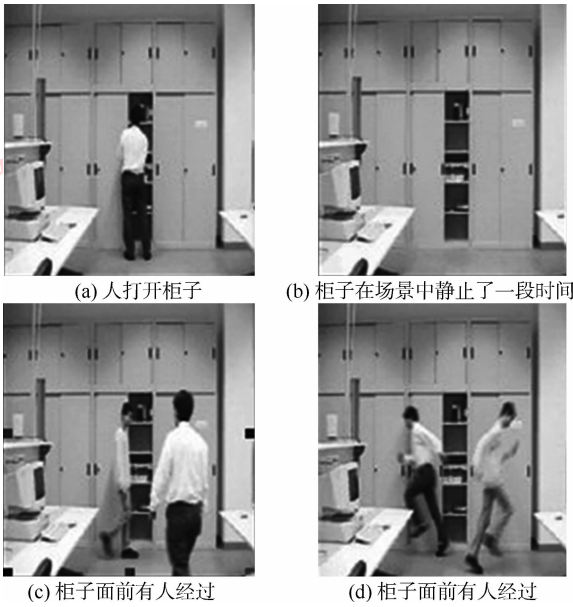


图 4 视频中的一帧
Fig. 4 A frame in the video

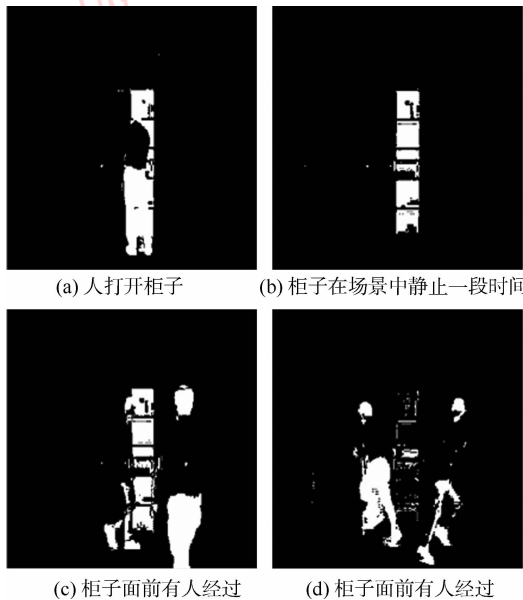


图 5 人体检测结果
Fig. 5 Human test results

出来,也就是说,背景减除法检测出来的前景包含运动人体和动态背景,而帧间差分法只能检测出运动人体。针对这个特点提出背景更新算法。为避免运动人体在场景中作短暂停留时,被更新到背景中这种情况,在背景更新算法中使用定时刷新策略,即给静止在场景中的物体或者人体一个时间期限,静止时间超过这个期限的物体或者人体则会被更新到背景中。背景更新算法描述如下:

if $|I_i - B_{i-1}| < M_0$ 则 $B_i = pI_i + (1 - p)B_{i-1}$
 else if $M_1/M_2 = 0$ 则
 if $|I_i - I_{i-1}| < M_3$ 则 $B_i = I_i$ 。

将被背景减除法检测为背景的像素按照一定的概率 p 更新到原来的背景中去;对被背景减除法被检测为前景的像素,判断帧数是否为 M_2 的倍数(M_2 为刷新的帧率),如果刚好为 M_2 的倍数,则进行刷新,具体的刷新操作是:如果某像素被帧间差分法检测为背景,则直接使用该像素的值替换原有的背景像素的值。

算法流程如下:

1) 从视频中得到一帧一帧图像,此时图像为 RGB 格式,将其分解为 R、G、B 3 个通道。选用 3×3 的高斯滤波器模板进行预处理。

2) 按照颜色空间转换式(1)~(3),得到 S、V、Cr 这 3 个通道的值。对这 3 个通道分别进行如下步骤 3)~5)的操作。

3) 利用视频的前 M_4 帧完成背景建模,即如果像素被帧间差分法检测出来为背景,则以一定概率更新到背景中去。

4) 利用背景减除法进行人体检测,同时进行背景更新。

5) 对人体检测的结果进行形态学处理,即先腐蚀后膨胀。

6) 对 S、V、Cr 这 3 个通道的检测结果进行融合,得到最终的检测结果。

2 实验结果

实验中选用的是在室内光照变化缓慢且基本无运动干扰的视频,视频分辨率为 320×240 ,帧数为 887 帧。部分参数设定如下:背景建模过程需要的帧数 M_4 ,取值为 50,更新速率 p 取值为 0.1,刷新背景帧数 M_2 为 20,即每隔 20 帧刷新一次。分别带入 5、10、15、20、25、30、35、40、45 这 9 个阈值,基于帧间差分法和背景减除法进行人体检测。对帧间差分法,Cb 通道的合适阈值为 5,其余 8 个通道均选择 25 作为阈值;对背景减除法,H 通道的合适阈值为 45,S 通道的合适阈值为 5~15,V 通道的合适阈值为 35,Cb 通道的合适阈值为 5,Cr 通道的合适阈值也为 5,其余通道的合适阈值均为 25。

原图如图 6 所示,图 7~10 均为实验结果,由图 6 可知,实验检测的人体的衣服颜色与背景颜

色相似,即存在伪装问题,则这次检测的重点难点部分为人体上身。对图 7—10 的图(a)进行研究:图 7、图 8 只能得到完整的头部和腿部,身子部分只有一点点轮廓;图 9 基本上只能得到完整的腿部;图 10 能够得到比较完整的头部、上身和腿部。对图 7—10 的图(b)进行研究;图 7、图 8 能够得到比较完整的头部;而图 9 能够得到部分身子和完整的腿部;图 10 能够得到完整的头部、腿部和身子。对图 7—10 的图(c)进行研究;图 7、图 8 只能得到完整的头部和腿部;身子部分只有一点点轮廓;图 9 基本上只能得到一个模糊的轮廓;图 10 能够得到比较完整的头部、上身和腿部。综上所述,实验提出的多颜色空间融合算法即使在伪装条件下也能够有效的检测出运动人体。

图 11 为图 4 对应的基于本文提出的背景更新算法的检测结果。在实验的视频中,第 312 帧也就是图 4(a)出现了人体打开柜子的动作,之后柜子一直处于打开状态,则柜子为动态背景。按照本文方

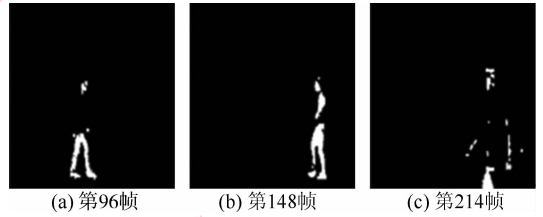


图 9 基于 YCrCb 的人体检测结果
Fig. 9 Human test results based on YCrCb

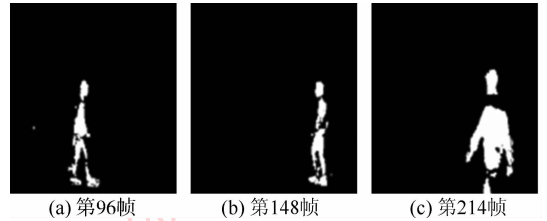


图 10 基于 SVCb 的人体检测结果
Fig. 10 Human test results based on SVCb

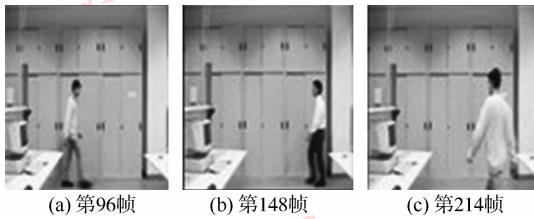


图 6 视频中的一帧
Fig. 6 A frame in the video

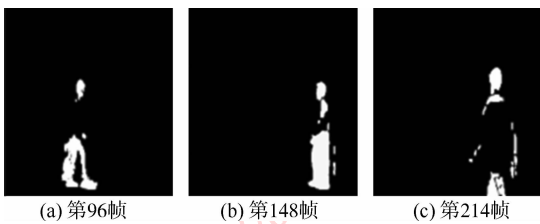


图 7 基于 RGB 的人体检测结果
Fig. 7 Human test results based on RGB

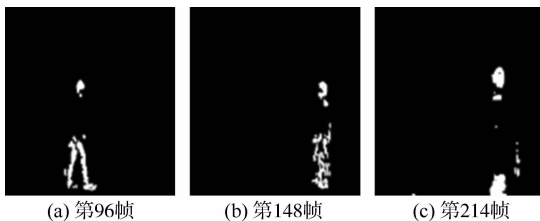


图 8 基于 HSV 的人体检测结果
Fig. 8 Human test results based on HSV

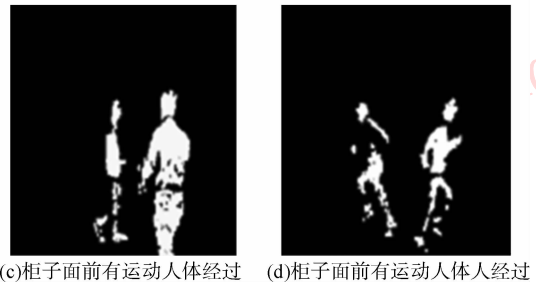
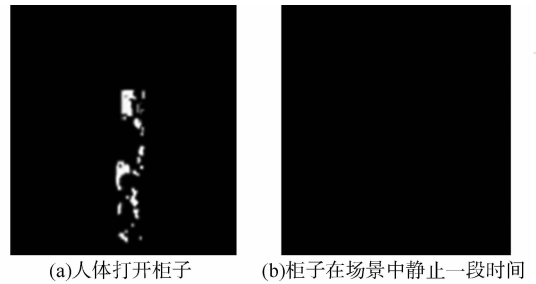


图 11 基于 SVCb 的人体检测结果
Fig. 11 Human test results based on SVCb

法,之后每隔 20 帧逐渐将动态背景更新为背景。在第 423 帧,如图 11(b)所示,打开的柜子这个动态背景已经完全融进了背景中,不再被检测出来。在第 468 帧,两个运动人体经过柜子边上,如图 11(c)所示,人体检测并没有受到动态背景的影响,仍然能够正常有效地检测出运动人体,第 685 帧,两个运动人体经过柜子边上如图 11(d)所示,人体检测并没有受到动态背景的影响,仍然能够正常有效地检测出运动人体。

3 结 论

针对摄像头固定、光线变化缓慢、存在伪装和动态背景的场景,提出基于颜色空间融合的人体检测算法,通过对实际图像序列测试可以看出,该算法得到的基于 SVCb 3 个颜色通道进行人体检测的检测效果要好于基于单个颜色空间的人体检测效果,即在伪装的情况下,能够更好地检测出运动人体。本文还提出了一种基于帧间差分法和定时刷新的背景更新方法,由实验结果可知,该背景更新算法能够很快地将场景中的动态背景更新到背景中去从而有效地避免对以后的运动人体检测造成影响。

参考文献 (References)

- [1] Heikkila M, Pietikainen M. A texture-based method for modeling the background and detecting moving objects [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(4) : 657-662.
- [2] Wu M J, Peng X R. Spatio-temporal context for codebook-based dynamic background subtraction [J]. International Journal of Electronics and Communications, 2010, 64(8) : 739-747.
- [3] Zha Y F, Bi D Y, Yang Y. Learning complex background by multi-scale discriminative model [J]. Pattern Recognition Letters, 2009, 30(11) : 1003-1014.
- [4] Liu Y Z, Yao H X, Gao W, et al. Nonparametric background generation [J]. Journal of Visual Communication & Image Representation, 2007, 18(3) : 253-263.
- [5] An G C, Zhang F J, Dai G Z. Mean shift using novel weight computation and model update [C] // Proceedings of IEEE ICASSP. Dallas: IEEE Press, 2010: 706-709.
- [6] Cheng J, Yang J, Zhou Y, et al. Flexible background mixture model for foreground segmentation [J]. Image and Vision Computing, 2006, 24(5) : 473-482.
- [7] Chen Y T, Chen C S, Huang C R, et al. Efficient hierarchical method for background subtraction [J]. Pattern Recognition, 2007, 40(10) : 2706-2715.
- [8] Wang H Z, Suter D. A consensus-based method for tracking : modelling background scenario and foreground appearance [J]. Pattern Recognition, 2007, 40(3) : 1091-1105.
- [9] Zhang X, Yang J. A novel algorithm to segment foreground from a similarly colored background [J]. International Journal of Electronics and Communications, 2009, 63(10) : 831-840.
- [10] Li Tongzhi, Ding Xiaoqing, Wang Shengjin. Human detection with a coarse-to-fine cascade linear SVM [J]. Journal of Image and Graphics, 2008, 13(3) : 566-570. [李同治, 丁晓青, 王生进. 利用级联 SVM 的人体检测方法 [J]. 中国图象图形学报, 2008, 13(3) : 566-570.]
- [11] Xu Cui, Zheng Ying, Wang Zengfu. Low-level image features based human body detection using hidden markov model [J]. Pattern Recognition and Artificial Intelligence, 2009, 22(5) : 743-749. [徐翠, 郑颖, 汪增福. 一种基于图像底层特征的隐马尔可夫人体检测方法 [J]. 模式识别与人工智能, 2009, 22(5) : 743-749.]
- [12] Song Xuehua, Chen Yu, Geng Jianfeng, et al. Moving object detection based on improved gaussian mixture background model [J]. Computer Engineering and Design, 2010, 31(21) : 4646-4649. [宋雪桦, 陈瑜, 耿剑锋, 等. 基于改进的混合高斯背景模型的运动目标检测 [J]. 计算机科学与设计, 2010, 31(21) : 4646-4649.]