面向演播环境的智能视觉信号处理方法研究

曾坤

(东南大学信息科学与工程学院；江苏 南京；211189)

摘　要:传统演播厅视频制作方式严重依赖于前期排练经验和摄像师的主观意识。本文通过图像处理算法对演播厅表演目标进行处理，提高了视频制作的智能性。首先使用混合高斯背景建模法实现了演播环境下的目标检测，并对检测目标进行了后处理，消除了目标的噪点和内部孔洞。然后使用 Meanshift 和 Kalman 滤波的结合算法实现了演播环境下的单目标、多目标跟踪。最后通过双目视觉技术提取到了演播厅演员目标的三维信息，完成了目标的实时有效定位。

关键词:目标检测；目标跟踪；双目视觉；三维信息

Research on Intelligent Visual Signal Processing for Studio Environment

ZengKun

(School of Information Science and Engineering,Southeast University; Nanjing China；211189)

Abstract:The traditional studio video production relies on the experience of rehearsals in the early stage and the subjective consciousness of the cameraman heavily. This paper deals with the performance targets in the studio through image processing algorithm, which improves the intelligence of video production. Firstly, the target detection in the studio environment is realized by using the mixed Gaussian background modeling method, and the detection target is post-processed to eliminate the noise points and internal holes of the target. Then, the single target and multi-target tracking is realized by using Meanshift and Kalman filtering algorithm Finally, the 3D information of the actor's target is extracted by binocular vision technology, and the real-time and effective target positioning is completed.

Key words: Target Detection; target tracking; 3D information extraction; binocular vision

演播环境下视频的录制，要追求艺术效果，充分展示人员（演员或主持）的肢体与面部细节，因此在正式表演之前要经过多次彩排才能找到最佳的拍摄角度，实际表演时还要限定表演者的运动路线，面部朝向等。这些过于死板的规则限制了表演者的发挥，影响视频效果。本文研究了一种新的演播方法，这种方法的基本思路是利用视觉图像处理技术，对表演者定位跟踪，提取其方位信息，将位置信息提供给主控单元，由主控单元控制摄像机完成自动拍摄。该方法能够减少视频制作对摄影师主观意识的依赖，提高拍摄效率和拍摄精度。

1 目标检测

本文使用混合高斯模型[1]来进行目标的背景建模。混合高斯模型可用如下形式表示：

$$p(X_t)=\sum_{i=1}^{K} w_{it} N(X_t,\mu_{it},\sum_{it}) \tag{1-1}$$

$$N(X_t,\mu_{it},\sum_{it})=\frac{1}{(2\pi)^{n/2}\left|\sum_{it}\right|^{1/2}}\times\exp(-\frac{1}{2}(X_t-\mu_{it})^T\sum_{it}^{-1}(X_t-\mu_{it})) \tag{1-2}$$

实际建模时通常 K=3～5个，$\mu_{it},\sum_{it}$ 为建立的第 i 个高斯模型对应的均值和协方差。

后面帧的像素点像素取值来临后，将按照下面的式(1-3)计算其与前面已有的所有高斯分布模型均值的距离：

$$|\mathrm{X}_t-\mu_{i,t-1}|\le c\delta_{i,t-1} \qquad (i=0,1,.....K) \tag{1-3}$$

若第 i 个高斯分布满足式(1-3)，则按照下式更新其对应的参数：

$$\begin{cases} w_{it}=\alpha(1-w_{it-1})+w_{it-1} \\ \mu_{it}=\beta(X_t-\mu_{it-1})+\mu_{it-1} \\ {\delta_{it}}^2=(1-\beta){\delta_{it-1}}^2+\beta(X_t-\mu_{it})^2 \end{cases} \tag{1-4}$$

其中 α、β 表示参数的更新速率。

引入均值滤波和形态学操作中的开操作对前面混合高斯背景建模法检测到的目标二值图像进行了进一步处理。使用 opencv 自带的标准视频测试了此目标检测算法。检测结果如图 1 所示。

(a) 标准视频图像　　(b) 检测结果

图 1 混合高斯背景建模及后处理后目标检测结果

2 目标跟踪

2.1 均值漂移算法

均值漂移算法[2]的基本思想是向样本空间数据点密度梯度上升的方向迭代计算均值漂移向量，最终迭代找到空间中样本分布最密集的区域。均值漂移向量示意图如下：

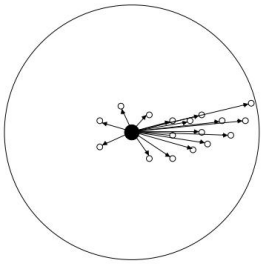


图 2 均值漂移向量示意图

设 S_h 是以 x 为中心点，半径为 h 的高维球围成的区域，这个球区域中有 k 个样本点，记作 $x_i(i=1,2.....k)$，引入核函数 $K(x)$ 后的均值漂移向量可以表示为：

$$M_h(x)=\frac{\sum_{x_i\in S_k} x_i k'\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{x_i\in S_k} k'\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}-x \tag{2-1}$$

其中 k' 是剖面函数 k 的导数，h 是核函数核窗宽度，每次迭代更新的圆心位置为：

$$x^{t+1}=\frac{\sum_{x_i\in S_k} x_i k'\left(\left\|\frac{x_t-x_i}{h}\right\|^2\right)}{\sum_{x_i\in S_k} k'\left(\left\|\frac{x_t-x_i}{h}\right\|^2\right)} \tag{2-2}$$

将当前均值漂移向量的起点移动到新圆心所在位置，并按照式(2-2)不断迭代计算下去，最终算法就会收敛到样本空间局部分布最集中的地方。

2.2 卡尔曼滤波算法

Kalman 滤波[3]是一种线性滤波器，其主要作用在包含不确定变化的动态系统中，利用系统的系列观测值对系统状态进行估计，且得到的这个估计值误差最小。Kalman 滤波中存在如下几个重要的方程：

$$\bar{x}_k = A\hat{x}_{k-1} \tag{2-3}$$

$$\hat{x}_k=\bar{x}_k+K_k(z_k-\bar{z}_k)=\bar{x}_k+K_k(z_k-H\bar{x}_k) \tag{2-4}$$

$$\bar{P}_k = AP_{k-1}A^T+Q \tag{2-5}$$

$$K_k=\bar{P}_kH^T(H\bar{P}_kH^T+R)^{-1} \tag{2-6}$$

$$P_k=(I-K_kH)\bar{P}_k \tag{2-7}$$

式中 z_k 为状态测量值，$\bar{x}_k$、$\bar{z}_k$ 分别表示忽略系统过程噪声和测量噪声后的系统状态预测值和测量值，$\hat{x}_k$ 为目标状态的后验状态估计值，Q、R 分别为过程噪声和测量噪声的协方差矩阵，K_k 是卡尔曼增益，P_k 和 $\bar{P}_k$ 分别是状态的先验和后验估计误差的协方差矩阵。式(2-3)、(2-5)实现了系统的先验估计，(2-4)实现系统的后验估计,(2-6)、(2-7)实现先验估计参数的更新。

2.3 结果测试

本文使用均值漂移算法和卡尔曼滤波算法的结合算法来实现目标跟踪，分别选取 CVRR_ATON 数据集、CDW_2012 数据集测试结合算法在单目标和多目标场景下的跟踪效果，算法运行结果如图 3 和图 4 所示（图片左上角为当前视频帧的编号），可以看到结合算法能够实现稳定的实时对象跟踪。

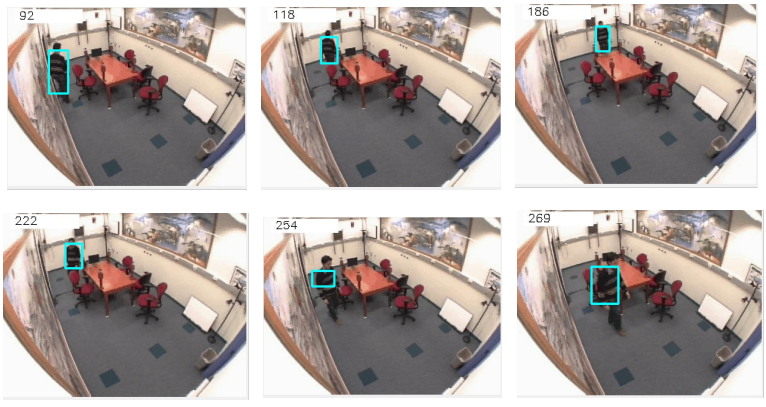


图 3 CVRR_ATON 标准数据集单目标跟踪测试结果

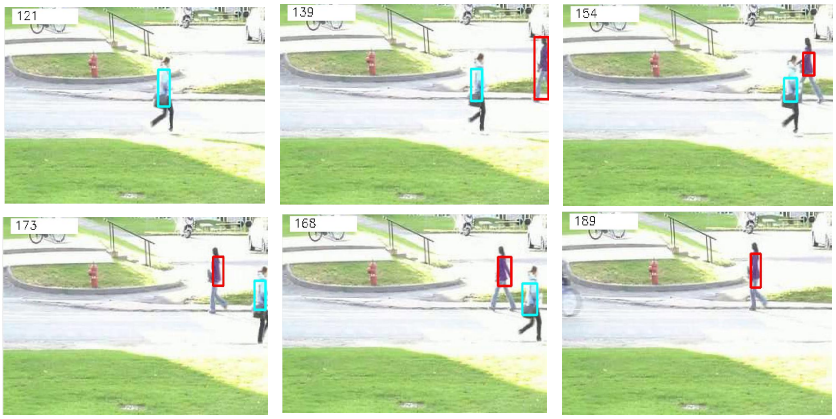


图 4 CDW_2012 标准数据集多目标跟踪测试结果

3 双目视觉定位

双目视觉是测量物体在场景中的深度信息或三维信息的一种常用技术，其基本原理是利用摆放在不同位置的两个相机采集待测目标物体的图像，通过目标物体在采集的两图像中存在的视差实现定位。

3.1 相机标定

相机标定简而言之就是一个确定相机几何参数并建立合理成像模型的过程。标定的质量将会大大影响目标定位、距离测量的精度。本文采用张正友标定法，使用 MATLAB2016a 自带的工具箱完成了标定。双目左右相机分别采集了 60 幅标定图片，初步标定后去除掉误差大的一些组，最终使用 28 组图片完成本次标定。最终的结果就是得到了左右两个相机各自的焦距、主点坐标以及反应两相机空间关系的旋转矩阵 R 和平移矩阵 T[4]，得到的具体相机参数如表 1 所示。

表 1 相机标定参数

参数	左相机	右相机
焦距 f_x (mm)	345.6519	340.7941
焦距 f_y (mm)	345.8345	340.6301
主点 u_0 (pixel)	319.0257	312.9203
主点 v_0 (pixel)	241.0603	229.8274
旋转矩阵 R	$\begin{bmatrix} 1 & 0.0040 & 0.0078 \\ -0.0040 & 1 & -0.0065 \\ -0.0078 & 0.0064 & 1 \end{bmatrix}$	
平移矩阵 T	$\begin{bmatrix} -63.18 & 0.0742 & 0.169 \end{bmatrix}$	

3.2 立体校正

由于双目视觉是根据空间中同一视点在不同位置相机上的成像位置差异，也即视差来实现定位的。那么必须要解决的问题就是左右相机同一点的匹配，也就是如何在右相机中搜索到左相机里的同一点。运用相机成像对极几何的知识来将横列方向的二维搜索转换为只进行横方向的一维搜索，可以有效的减少匹配的搜索量，这也正是立体校正的目的。

3.3 立体匹配

本文使用 SGBM 立体匹配算法，该算法主要包含预处理、代价计算、视差优化、后处理四个步骤。SGBM 算法得到代价后会建立作用于整个视差图像的能量函数，用于优化生成的视差图效果。能量函数[5]的表达式为：

$$E(D)=\sum_{p}(C(p,D_p)+\sum_{q\in N_p}P_1T\left[\left|D_p-D_q\right|=1\right]+\sum_{q\in N_p}P_2T\left[\left|D_p-D_q\right|>1\right) \tag{3-1}$$

式中，D 表示视差图，$C(p,D_p)$ 表示像素点 p 视差值为 D_p 时计算得到的匹配代价，N_P 表示 p 点周围 8 连通区域的像素点。P_1 为邻近像素点的视差变化值等于 1 时的惩罚因子，P_2 为邻近像素点的视差改变值比 1 大时的惩罚因子，且要求 P_2 一定要大于 P_1 。$T[.]$ 如果函数为真则返回 1，否则返回 0。

选取 Middlebury 数据集提供的图片测试 SGBM 算法的匹配效果，最终生成的视差图如图 5 所示。

(a) (b) (c)

图 5 立体匹配测试结果 (a) 左图原始图像 (b)右图原始图像 (c) SGBM 算法生成的视差图

3.4 三维重建

立体校正这一步骤完成后，除了得到使得极线行对准的左右相机变换矩阵外，还将得到它们分别的投影矩阵 p_r,p_l，以及一个重投影矩阵 Q。由这个矩阵便可以建立二维像素坐标和视差值与三维世界坐标的联系，Q 的具体形式如下：

$$Q=\begin{bmatrix}1 & 0 & 0 & -c_x\\ 0 & 1 & 0 & -c_y\\ 0 & 0 & 0 & f\\ 0 & 0 & \frac{-1}{T_x} & \frac{c_x-c_x'}{T_x}\end{bmatrix} \tag{3-2}$$

其中，f 是左相机焦距，T_x 是两个相机投影中心 X 方向的水平距离，c_x、c_x' 分别为主点在所采集的左右图像上的横坐标[6]。有了投影矩阵便可以借助前面生成的场景视差值将某个像素点 (x,y) 映射到三维，映射关系如下：

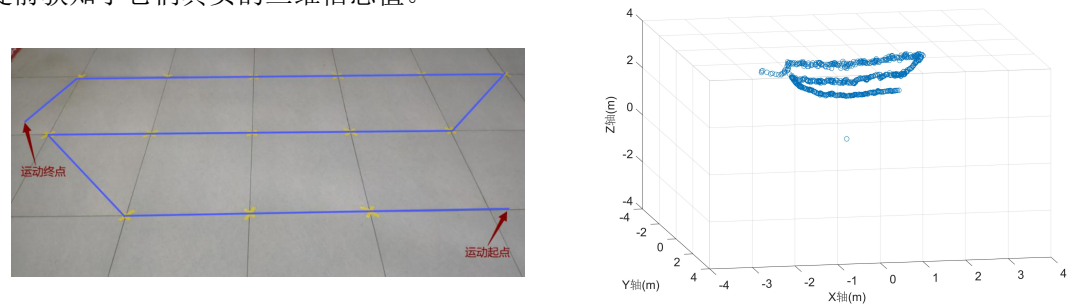
$$Q\begin{bmatrix}x\\ y\\ d\\ 1\end{bmatrix}=\begin{bmatrix}x-c_x\\ y-c_y\\ f\\ \frac{-d+c_x-c_x'}{T_x}\end{bmatrix}=\begin{bmatrix}X\\ Y\\ Z\\ W\end{bmatrix} \tag{3-3}$$

其中 d 表示其视差值，设该点在场景中的三维坐标是 $(\frac{X}{W},\frac{Y}{W},\frac{Z}{W})$，（W 一般为 1）然后可以进一步求解到：

$$\frac{X}{W}=\frac{-T_x(x-c_x)}{d-(c_x-c_x')}\qquad \frac{Y}{W}=\frac{-T_x(y-c_y)}{d-(c_x-c_x')}\qquad \frac{Z}{W}=\frac{-T_xf}{d-(c_x-c_x')} \tag{3-4}$$

3.5 定位测试

本文利用室内环境模拟演播厅录制了人体目标运动视频。约定目标按照预定轨迹以一定速度运动，利用前面小节实现的目标跟踪算法对目标进行了实时跟踪，再利用本小节的双目视觉技术对跟踪到的每一帧目标进行了定位。为了准确测量定位误差，实验开始前已经在室内场景中标注了 14 个特征点，并且提前获知了它们真实的三维信息值。



(a)目标真实运动轨迹　　(b)双目视觉定位得到的运动轨迹

图 6 人体运动目标真实轨迹和测量轨迹对比图

图 6 中给出了目标真实运动轨迹和测量轨迹的对比图，从中可以看到通过双目视觉定位技术生成的测量轨迹和预设的真实轨迹十分吻合，实现了目标定位的目的。为了定量说明双目视觉定位的精度，本文又对标志的 14 个特征点的测量三维数据进行了统计，并与它们各自真实的三维数据进行了对比，最终绘制了图 7 所示的特征点三维数据、距离测量的百分比误差图。

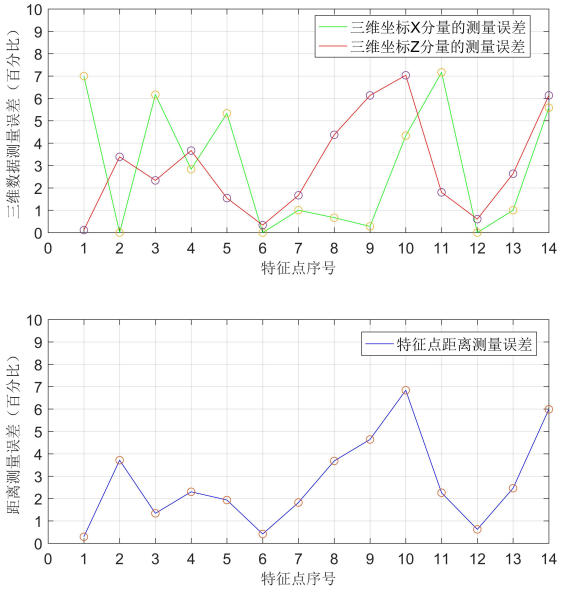


图 7 特征点三维数据与距离测量百分比误差

由图 7 可以看到，三维坐标和距离的最大测量误差都在 7%左右，其中测距误差比较大的点是距离相机最远的特征点 9、10 和 14，而距离较近的那些特征点，测量误差全部低于 4%，定位效果达到了预期目标。同时这也进一步印证了本文采用的目标检测、跟踪算法确实具有优异的性能。将这些数据上传给主控单元，即可实现演播环境下的智能视频制作。

参考文献

[1] 罗超宇,李小曼,韩骏浩, 等.改进的混合高斯背景建模算法[J].计算机应用与软件,2015,(10):209-212,230.
[2] 沈豪,庄建军,郑茜颖, 等.一种基于 MeanShift 算法的目标跟踪系统的设计与实现[J].电子测量技术,2018,41(14):11-15.
[3] 赵广辉,卓松,徐晓龙.基于卡尔曼滤波的多目标跟踪方法[J].计算机科学,2018,45(8):253-257,276.
[4] 迟德霞,王洋,宁立群, 等.张正友法的摄像机标定试验[J].中国农机化学报,2015,36(2):287-289,337.
[5] 张欢,安利,张强, 等.SGBM 算法与 BM 算法分析研究[J].测绘与空间地理信息,2016,(10):214-216.
[6] 杨晨曦,华云松.基于双目立体视觉的目标物测距研究[J].软件,2020,41(1):128-132.