doi:10.3788/gzxb20124104.0461

基于改进 Mean Shift 算法的实时视频目标跟踪

胡波[1,2]

(1 浙江东方职业技术学院 工程技术系，浙江 温州 325011)

(2 宁波大学 信息科学与工程学院，浙江 宁波 315211)

摘　要:设计了一套嵌入式平台上实现的视频目标跟踪系统. 该系统采用 CMOS 图像传感器获取视频信号，利用 Z228 多媒体芯片自带的 ARM9 处理器完成视频信号的控制，并通过 MPEG-4 硬件编码器实现视频信号的压缩. 用 Mean Shift 算法跟踪运动目标，针对其收敛的局限性设置多个搜索点来提高其跟踪效果. 通过减少采样点和标记已计算点来提高代码运行速度，增强了跟踪的实时性. 实验结果表明，本系统能以 27 fps 速率连续稳定地实现视频目标的跟踪.

关键词:嵌入式;视频目标跟踪;图像传感器;Mean Shift 算法;实时

中图分类号:TP391　　**文献标识码**:A　　**文章编号**:1004-4213(2012)04-0461-5

0　引言

智能视频监控是当前计算机视觉研究领域的热点问题. 基于视频序列的运动目标跟踪技术则是智能视频监控的基础，它为视频行为理解、场景分析等很多后续处理提供了基础. 人们针对不同的应用环境提出了许多跟踪算法，如基于粒子滤波[1]、Kalman 滤波[2]、extended Kalman 滤波[3]、Unscented Kalman 滤波[4]、Mean Shift 算法[5-6]等. 但是目前视频目标跟踪算法大多基于 PC 机实现，不利于在实际应用中推广. 视频目标跟踪算法需要处理的数据量大、运算复杂，需要性能强大的处理器才能实时处理. 嵌入式视频处理系统[7-9]利用嵌入式硬件平台一体化的优势，实现视频信号的采集、传输、处理，具有实时性好、体积小、功耗低等优点. 本文设计的系统选用 Z228 多媒体应用处理器作为嵌入式开发平台，用 CMOS 图像传感器[10-11]获取视频信号，对传统 Mean Shift 跟踪算法进行优化，实现了视频目标的实时跟踪.

1　系统设计

1.1　Z228 处理器

本文设计的系统选用的 Z228 多媒体应用处理器是一款 0.13 μm 工艺的大规模片上系统芯片，以 ARM926EJ 和 MPEG-4 硬件全双工编解码器为核心. ARM926EJ 的主频达到 266 MHz，包含一个 16 KB的高速指令缓存(ICache)和一个 16 KB 高速数据缓存(DCache). 处理器的内存管理单元(MMU)支持 WindowsCE、Linux 和 Symbian 等嵌入式操作系统. Z228 内置 MPEG-4 硬件编解码引擎，在该硬件编码器支持下，Z228 以 30 fps 的速率处理 VGA 信号时占有内核的负荷不到 5%，可以流畅运行操作系统和其它应用程序. Z228 处理器拥有强大的视频图像处理功能，主要包括摄像头输入、TV 信号输入、TV 信号输出、MPEG-4 编码和 MPEG-4 解码这 5 个功能，结构和信号传输如图 1.

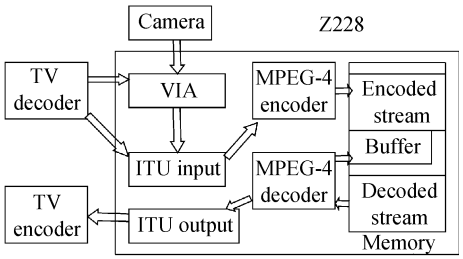

图 1　Z228 处理器视频处理模块结构

Fig. 1　Structure diagram of video processing module for Z228

1.2　CMOS 图像传感器

Z228 通过 VIA(视频接口加速器)模块来控制图像传感器，工作模式有:视频采集、拍照、预览和空闲四种. 本系统中 VIA 工作在视频采集模式，设置控制寄存器 VIA_RAWPRV_CONTROL 的模式控制位(字节低 2 位)为“11”，并设置启停控制位(字节第 3 位)为 1. 此时，VIA 模块将持续捕获传感器的视频数据，经过 VIP 子模块处理后送到 MPEG-4 编码器进行压缩并送液晶显示器显示. 视频采集模式

基金项目:国家自然科学基金(No. 60971074)资助

第一作者:胡波(1982－)，男，讲师，硕士，主要研究方向为视频信号处理与嵌入式系统. Email: hubo9511@163.com

收稿日期:2011-07-13;**修回日期**:2011-09-06

输出 YUV 格式图像数据，该数据最终存放在 FIFO 中并通过 DMA 的方式传送到存储器中. Z228 处理器与 OV9655 传感器硬件连接如图 2. 图中处理器的 CSI 接口与 OV9655 连接，完成视频信号的传输控制. Z228 利用 I^2C 总线对 OV9655 内部寄存器进行编程，设置其工作方式和数据输出格式等内容. 本系统中，在 OV9655 和 Z228 处理器之间增加 SN74ALVCH16245 芯片来提高总线驱动能力. 同时通过将外部摄像头同底板进行隔离的方法来提高系统的可靠性.

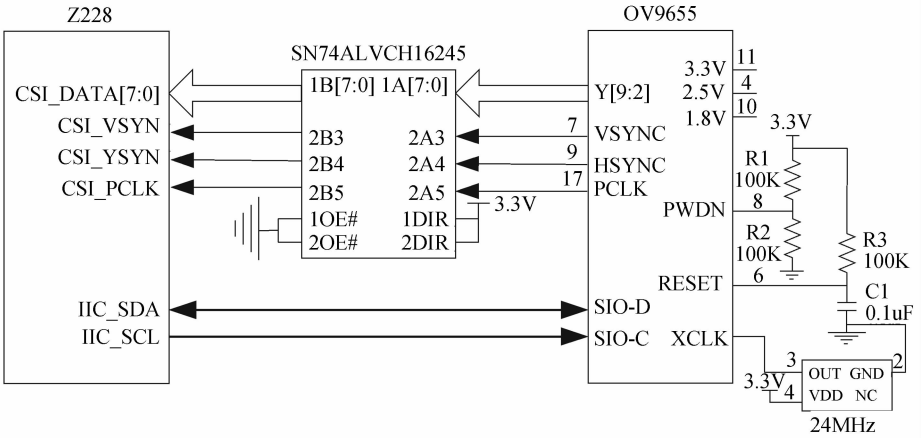

图 2 OV9655 图像传感器硬件连接图

Fig. 2 Hardware connection diagram of OV9655 image sensor

2 视频目标跟踪算法

2.1 初始帧目标特征模型建立

首先在初始帧中确定跟踪目标，并将其转化为 HSV 空间的 H 分量图直方图，如图 3(b). 在该直方图中，将每个灰度值用其所代表的频率数同等级的灰度值来表示，得到概率投影图，如图 3(c). 在初始帧中第 u 个特征概率由式(1)定义

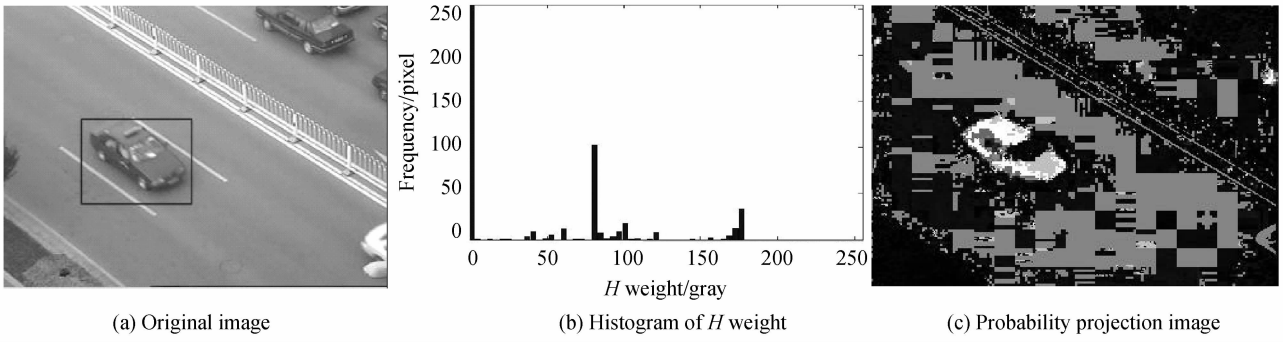

图 3 原图、颜色分量直方图及概率投影图

Fig. 3 Original image and histogram of H weight and probability projection image

$$\hat{q}_u = C\sum_{i=1}^{n} k\left(\left\|\frac{t_i - d_0}{h}\right\|^2\right)\delta[b(t_i) - u] \tag{1}$$

式中 d_0 表示搜索窗口的中心坐标；$t_i(t_{ix}, t_{iy})$ 表示第 i 个像素的坐标；n 表示像素总个数；$k(\|x\|^2)$ 表示核函数，系统采用 Epanechnikov 核函数以减少计算量；h 表示核函数的带宽；函数 δ 和 b 表示特征中第 i 个像素代表的颜色值；C 是满足 $\sum_{u=1}^{m}\hat{q}_u = 1$ 的归一化系数，其中 m 表示特征个数.

2.2 当前帧目标特征模型建立

在当前帧中，第 u 个特征概率由式(2)定义

$$\hat{p}_u(d_1) = C_h\sum_{i=1}^{n_h} k\left(\left\|\frac{t_i - d_1}{h}\right\|^2\right)\delta[b(t_i) - u] \tag{2}$$

式中 $d_1(d_{1x}, d_{1y})$ 表示当前帧搜索窗口的中心坐标.

2.3 特征匹配规则

通过定义相似性函数来衡量初始帧目标和当前帧目标的匹配程度，相似性函数为

$$\hat{\rho}(t) \equiv \rho\left[\widehat{p(t)}, \hat{q}\right] = \sum_{u=1}^{m}\sqrt{\hat{p}_u(t)\hat{q}_u} \tag{3}$$

传统 Mean Shift 以前一帧目标的位置作为当前帧的起始搜索点，在该点领域内寻找目标的最优位置 d_c. 将式(3)在 d_1 处泰勒展开，取其线性部分

$$\rho\left[\widehat{p(t)}, \hat{q}\right] \approx \frac{1}{2}\sum_{u=1}^{m}\sqrt{\hat{p}(d_1)\hat{q}_u} + \frac{C_h}{2}\sum_{i=1}^{n_h} w_i k\left(\left\|\frac{t - t_i}{h}\right\|^2\right) \tag{4}$$

式中加权系数 $w_i = \sum_{u=1}^{m}\delta[b(t_i) - u]\sqrt{\frac{\hat{q}_u}{\hat{p}_u(d_1)}}$.

通过式(4)求极值，就能够完成目标最优匹配. 将式(4)求导即得 Mean Shift 向量

$$M_{h,G}(t)=d_c-d_1=\frac{\sum_{i=1}^{n_h}G\left(\left\|\frac{t_i-d_1}{h}\right\|^2\right)w_i t_i}{\sum_{i=1}^{n_h}G\left(\left\|\frac{t_i-d_1}{h}\right\|^2\right)w_i}-d_1 \quad (5)$$

由于核函数选用 Epanechnikov 函数，式中

$$G(x)=\begin{cases}1 & \text{if} \quad \|x\|<1\\ 0 & \text{if} \quad \|x\|\geqslant 1\end{cases},$$

因此 d_c 就是简单的加权计算

$$d_c=\sum_{i=1}^{n_h}w_i t_i/\sum_{i=1}^{n_h}w_i,$$

通过连续地迭代计算 $d_1\leftarrow d_c$，d_c 将会收敛于局部极大值，即在当前帧目标达到最优位置.

2.4　迭代误差处理

在当前帧计算目标的相似度时，若满足 $\rho[\widehat{p(d_c)},\hat{q}]>\rho[\widehat{p(d_1)},\hat{q}]$，则以 d_c 为新的搜索点重新计算，直到满足 $\|d_c-d_1\|<\varepsilon$ 为止. 若在计算中引入了飘移误差，则会出现 $\rho[\widehat{p(d_c)},\hat{q}]<\rho[\widehat{p(d_1)},\hat{q}]$ 情况. 令 $d_c=(d_1+d_c)/2$，代入计算相似度，使满足 $\rho[\widehat{p(d_c)},\hat{q}]>\rho[\widehat{p(d_1)},\hat{q}]$，则可以避免由于漂移误差引起的无效计算. 误差 ε 设置为小于一个像素的值，在实际的跟踪过程中迭代计算次数一般少于 10 次.

2.5　算法改进

Mean Shift 算法的时间复杂度和其向量的迭代次数有关，为了进一步提高跟踪实时性，通过减少采样点和标记已计算的搜索点对算法进行优化. 同时通过适当扩大搜索范围的方法来提高跟踪精确性.

2.5.1　目标采样

Mean Shift 算法采用颜色直方图作为目标匹配的特征，与目标尺寸的大小无关. 考虑在目标较大时，对其按照一定的规则采样以后再进行直方图的计算，从而减小每次迭代时间，提高了跟踪的实时性. 当目标长度 L_x 较大时，设大于 $L_{x\max}$（$L_{x\max}$ 根据实际场景中目标的大小确定）时，x 方向上采样的横坐标为

$$x=d_{1x}+\frac{i}{2}\times\frac{L_x}{L_{x\max}} \quad (i=0,\pm1,\pm2\cdots\pm L_{x\max}) \quad (6)$$

当目标高度较大时，在 y 方向上做同样的采样. 通过实验表明该采样模式对匹配准确性影响很小，但处理速度可以明显改善.

2.5.2　标记已计算过的点

在同一个搜索区域中有多个搜索起始点，Mean Shift 算法通过迭代运算使很多点都会收敛于同一点上，因此会出现重复运算. 改进算法通过对已经计算过的点进行标记，一旦碰到标记点则停止搜索，避免重复运算.

2.5.3　扩大搜索范围

由于 Mean Shift 算法在跟踪目标时，必须保证上一帧的运动目标不能完全脱离当前帧目标的位置，否则 Mean Shift 算法就会跟踪不上目标. 事实上当运动目标速度较快时，目标在连续两帧的图像空间上完全有可能不重叠. Mean Shift 算法在上一帧目标的位置区域内迭代求解取得局部最优值，而得不到当前帧中目标实际位置，导致跟踪失效. 因此在跟踪过程中通过设置多个搜索点来适当扩大搜索范围，搜索区域中心位置表示为

$$\begin{cases}d'_{1x}=d_{1x}+\frac{i}{2}\times L_x & (i=0,\pm1,\pm2\cdots\pm N_x)\\ d'_{1y}=d_{1y}+\frac{j}{2}\times H_y & (j=0,\pm1,\pm2\cdots\pm N_y)\end{cases} \quad (7)$$

式中 H_y 为目标的高度，$N_x\times N_y$ 表示搜索范围，由式(8)定义

$$N_x=\mathrm{ceil}\left(\frac{2V_x\times\Delta t}{L_x}\right);N_y=\mathrm{ceil}\left(\frac{2V_y\times\Delta t}{L_y}\right) \quad (8)$$

式(8)中 ceil()表示取整函数，Δt 表示相邻视频帧的时间间隔，V_x、V_y 表示当前帧中运动目标在水平和垂直方向的速度估计值，定义为

$$V_x=\frac{|d'_{cx}-d_{1x}|}{\Delta t};V_y=\frac{|d'_{cy}-d_{1y}|}{\Delta t} \quad (9)$$

式中 $d'_c(d'_{cx},d'_{cy})$ 表示上一帧计算得到的运动目标最优位置.

3　实验结果及分析

为了验证本系统的跟踪效果，在实验室环境下对人体运动目标进行测试研究. 图 4 为本文研究方法，在初始帧中检测运动目标[12]并用矩形框来描述. 此时目标处于部分遮挡状态，跟踪算法提取了人体的衣服及头部特征. 在第 132 帧，目标已经完全脱离遮挡状态，由于 Mean Shift 算法的跟踪窗口是固定不变的，所以该算法在当前帧中搜索一个和初始帧目标最相似的窗口作为跟踪窗口. 在第 256 帧，目标又开始处于部分遮挡状态，此时 Mean Shift 算法仍然能够根据衣服纹理特征跟踪目标，直至目标完全消失. 若直接采用 Mean Shift 算法跟踪效果如图 5. 第 128 帧以后运动目标速度加快，而传统 Mean Shift 算法是在上一帧目标所在位置区域内迭代求取局部最大值，和目标实际位置偏离很大，导致了跟踪失败.

(a) 30th frame (b) 132nd frame (c) 178th frame (d) 256th frame

图 4 改进 Mean Shift 跟踪效果

Fig. 4 Tracking results with improved mean shift

(a) 45th frame (b) 102nd frame (c) 128th frame (d) 175th frame

图 5 Mean Shift 跟踪效果

Fig. 5 Tracking results with mean shift

图 6 是人脸跟踪效果. 从图中可以发现,当目标发生旋转(321 帧)和部分遮挡时(66 帧),本文方法仍然能够很好地实现跟踪. 采用 Mean Shift 算法及其改进算法对部分帧处理时间如表 1. 从表中可以发现,Mean Shift 算法经过采样或标记后,对每帧图像处理速度均有提高(平均处理时间由 45 ms 变为 39 ms 和 39 ms),同时采用采样和标记后处理速度明显提高(平均处理时间为 34 ms). 适当扩大搜索范围后,处理速度略有下降(平均处理时间为 37 ms),但并不影响跟踪的实时性.

(a) 15th frame (b) 66th frame (c) 189th frame (d) 251th frame

(e) 321st frame (f) 366th frame (g) 489th frame (h) 551st frame

图 6 目标旋转、部分遮挡跟踪

Fig. 6 Tracking results for object rotation and partial occlusion

表 1 处理时间对比 (毫秒/帧)

Tabel 1 Comparison of processing time (ms/frame)

Frame	15	68	113	156	189	221	251	286	322	371	402	456	501	550
Mean Shift	46	40	43	47	44	49	44	44	46	44	49	46	46	42
Sampling	44	37	39	40	39	41	38	37	39	40	41	39	40	37
Mark	43	35	40	39	37	42	40	39	38	37	39	42	39	36
Sampling and mark	34	31	35	36	31	35	33	34	36	34	35	32	37	34
Proposed method	37	34	37	38	35	38	36	36	38	37	38	35	39	36

4 结论

本文设计了一种采用 Z228 多媒体处理芯片和标准 CMOS 图像传感器结构的视频目标跟踪系统. Z228 多媒体处理芯片提供了 ARM9 嵌入式开发平台,同时其自带的 MPEG-4 硬件编码器能够对视频

信号进行实时压缩和传输. 设计了 Z228 和 OV9655 图像传感器的接口电路，保证其能够稳定可靠地工作. 在 EVC 集成开发环境中实现目标跟踪算法. 通过对 Mean Shift 算法的优化提高了跟踪的实时性，同时增强了其跟踪的有效性. Z228 硬件资源丰富，下一步的工作可以考虑增强其网络传输功能，同时进一步增强其智能分析能力，开发一套智能的网络摄像机.

参考文献

[1] WANG Zhao-wen, YANG Xiao-kang, XU Yi, *et al*. Camshift guided particle filter for visual tracking [J]. *Pattern Recognition Letters*, 2009, **30**(4): 407-413.

[2] WENG Shi-ku, KUO Chung-ming, TU Shu-kang. Video object tracking using adaptive Kalman filter[J]. *Journal of Visual Communication and Image Representation*, 2006, **17**(6): 1196-1197.

[3] DONG Zhe, YOU Zheng. A novel extended Kalman filter for a class of nonlinear systems[J]. *Progress in Natural Science*, 2006, **16**(9): 912-918.

[4] LI Pei-hua, ZHANG Tian-wen, MA Bo. Unscented Kalman filter for visual curve tracking [J]. *Image and Vision Computing*, 2004, **22**(2): 157-164.

[5] HU Bo, CHEN Ken, XU Jian-yu, *et al*. Video object tracking based on Mean-shift with self- adaptability search window and Kalman prediction[J]. *Journal of Optoelectronics · Laster*, 2009, **20**(11): 1517-1522.
胡波，陈恳，徐建瑜，等. 基于 Kalman 预测和 Mean-shift 算法的视频目标跟踪[J]. 光电子·激光，2009，**20**(11)：1517-1522.

[6] COMANICIU D, MEER P. Mean Shift: a robust application toward feature space analysis [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(5): 603-619.

[7] BRAMBERGER M, DOBLANDER A, MAIER A, *et al*. Distributed embedded smart cameras for surveillance applications[J]. *Computer*, 2006, **39**(2): 68-75.

[8] CHATTOPADHYAY A, BOULT T E. PrivacyCam: a privacy preserving camera using uCLinux on the blackfin DSP [C]. Minneapolis: In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2007, 1-8.

[9] SCHIFF J, MEINGAST M, MULLIGAN D K, *et al*. Respectful cameras: detecting visual markers in real-time to address privacy concerns[C]. San Diego: In Proceedings of the IEEE International Conference on Intelligent Robots and Systems, 2007, 971-978.

[10] ZHAO Zhi-gang, DU Yang, HUANG Jian-heng, *et al*. Implementation of a CMOS imaging system with adjustable power supplies and driving timing [J]. *Opto-Electronic Engineering*, 2010, **37**(6): 119-125.
赵志刚，杜杨，黄建衡，等. 电源和驱动时序可调整 CMOS 成像系统的实现[J]. 光电工程，2010，**37**(6)：119-125.

[11] PENG Xiao-qi, SUN Yuan. Filtering and temperature correction algorithms for smog interference in temperature measurement based on CCD image sensor[C]. Zhangjiajie: In Proceedings of the IEEE Computer Society Conference on Measuring Technology and Mechatronics Automation, 2009, 24-27.

[12] QIN Jian, CHEN Qian, QIAN Wei-xian. A detection algorithm for dim and small infrared target based on the optical flow estimation and the adaptive background suppression[J]. *Acta Photonica Sinica*, 2011, **40**(3): 476-482.
秦剑，陈钱，钱惟贤. 基于光流估计和自适应背景抑制的弱小目标检测[J]. 光子学报，2011，**40**(3)：476-482.

Real-time Video Target Tracking Based on Improved Mean Shift Arithmetic

HU Bo[1,2]

(1 *Department of Engineering Technology*, *Zhejiang Dongfang Vocational and Technical College*, *Wenzhou*, *Zhejiang* 325011, *China*)

(2 *Faculty of Information Science and Engineering*, *Ningbo University*, *Ningbo*, *Zhejiang* 315211, *China*)

Abstract: A video target tracking system is designed which is achieved in the embedded platform. The system acquires video signal by using CMOS image sensor, controls video signal in ARM9 processor which is obtained in Z228 multimedia chip, and compresses video signal through MPEG-4 hardware encoder. Meanwhile, the Mean Shift arithmetic is used to track moving target. Considering the convergence limitation, multiple search point is set to improve the tracking results. The operating speed of code is raised through decreasing sampling point and labeling computed points, thus the tracking instantaneity is enhanced. The experiment results indicate that this system can achieve video target tracking continually and steadily at the speed is 27 fps.

Key words: Embedded; Video target tracking; Image sensor; Mean Shift arithmetic; Real-time