

基于多特征 Mean Shift 的人脸跟踪算法

张涛 蔡灿辉

(华侨大学信息科学与工程学院 泉州 362021)

摘要: 该文把局部三值模式(Local Ternary Patterns, LTP)纹理特征引入 Mean Shift 跟踪算法, 提出了基于多特征的 Mean Shift 人脸跟踪算法以解决 Mean shift 跟踪算法的鲁棒性问题。通过对 LTP 纹理特征的分析、研究, 提出了一个 LTP 关键纹理模型, 既增强了目标的关键纹理信息, 又简化了 LTP 纹理模型。在此基础上, 提出一种基于 LTP 关键纹理特征和肤色特征的 Mean Shift 人脸跟踪算法, 有效地解决了 Mean Shift 算法的鲁棒性问题。为进一步提高对快速运动目标的跟踪速度和跟踪性能, 该文引入了卡尔曼滤波器对目标进行预测。实验结果表明, 该文的算法在目标定位的准确性和跟踪性能上比 Mean Shift 算法均有明显的提高。

关键词: 人脸跟踪; 卡尔曼滤波; Mean Shift 算法; 局部三值模式

中图分类号: TP391.41

文献标识码: A

文章编号: 1009-5896(2009)08-1816-05

A Face Tracking Algorithm Based on Multiple Feature Mean Shift

Zhang Tao Cai Can-hui

(Institute of Information Science and Technology, Huaqiao University, Quanzhou 362021, China)

Abstract: In this paper, an improved Mean Shift face tracking algorithm based on Local Ternary Patterns (LTP) of texture and color features is proposed to improve the robustness of the Mean Shift algorithm. Based on the study of LTP texture features, an LTP key texture pattern is introduced to enhance the important features of an object and reduce the computational complexity of the LPT texture model. A multiple feature Mean Shift face tracking algorithm is then proposed based on the LTP key texture and complexion features, and the robustness of Mean Shift algorithm is significantly enhanced. Furthermore, in order to improve the tracking speed and robustness, the Kalman filter is introduced to predict the position of the object window. Experimental results show that compared with the original Mean Shift algorithm, the proposed multiple feature face tracking algorithm has significantly improved the tracking performance.

Key words: Face tracking; Kalman filter; Mean Shift algorithm; Local Ternary Patterns (LTP)

1 引言

鲁棒的实时人脸跟踪既是智能视频监控和基于视觉的人机交互以及机器人导航等领域的一个核心问题, 又是视频分析与理解的基础。近年来研究者们已经提出了许多跟踪算法, 其中 Mean Shift^[1]及其改进算法具有计算简单, 实时性好等特点, 得到了广泛的应用。然而, 作为一种基于核密度估计的无参快速模式匹配算法, Mean Shift 能否准确地搜索到跟踪对象取决于目标模式的表示方法。现有的 Mean Shift 及其改进算法通常采用颜色特征, 在实际应用环境中, 由于场景变化较大, 很难保证跟踪算法的鲁棒性。当亮度变化过于剧烈或跟踪目标和背景颜色相近时, 容易丢失目标, 导致跟踪失败。

因此, 不少文献在 Mean Shift 算法中引入多特征目标模型^[2-4], 提高跟踪算法的鲁棒性, 但它们大多是基于贝叶斯概率框架, 因而计算复杂度较高, 难于达到跟踪的实时性。

考虑到目标的纹理通常不会受光照和背景颜色影响, 是较为稳定的特征, 因此, 把纹理特征融合到目标模型来表示目标是一个更好的选择。常用的纹理模型(如灰度共生矩阵, 小波纹理模型等)虽然能较好地描述目标的纹理特征, 但由于它们不是点样本估计, 难于融合颜色等其它特征。而 Ojala 等^[5]提出的 LBP(Local Binary Patterns)纹理模型与传统的颜色模型一样, 属于点样本估计方式, 并且具有尺度不变性、旋转不变性和计算复杂度低等优点, 成为当前基于纹理算法研究中的一个热点。Nguyen 等^[6]把 LBP 简单纹理模型与图像灰度模型相结合应用于视频跟踪。通过手动提取运动目标的纹理直方图和灰度直方图组成二维目标模板, 并将该模板嵌

2008-09-04 收到, 2009-03-09 改回

国家自然科学基金(60772164), 福建省自然科学基金(A0710009)和福建省科技计划项目(2005H034)资助课题

入到Mean Shift跟踪算法中,该方法对单色温度录像仪所录的视频序列具有较好的跟踪效果。但若目标灰度受噪声干扰或与背景难以区分时,容易丢失目标。Wang^[7]等把目标的颜色特征和LBP简单纹理特征分别作为Mean Shift跟踪算法的模板,并利用Fisher准则对两种特征进行可分性分析,以调节各个模板的系数。该方法能对单目标稳定跟踪,但计算复杂度大,实时性较差。Ning等^[8]通过分析LBP纹理模型的各个模式,选取LBP纹理模型中的5个关键模式与红色分量、绿色分量组成三维目标模型嵌入到Mean Shift跟踪算法中。该方法具有很好的实时性,但容易受噪声干扰,并且由于舍弃了较多的LBP纹理模式,容易造成目标定位不准。

考虑到LBP纹理模型过于依赖中心点像素的灰度值,对噪声比较敏感,Tan等^[9]提出了LTP(Local Ternary Patterns)纹理模型,有效地克服了LBP对噪声比较敏感的缺点,并成功应用于人脸识别中。本文以LTP纹理模型为研究对象,通过对LTP纹理模型各种模式的分析,选取关键模式构建简化的LTP纹理模型来表达目标的纹理特征,并结合颜色特征构成多特征目标函数嵌入到Mean Shift跟踪算法中,以提高Mean Shift跟踪算法的可靠性。通过引入卡尔曼滤波器对目标位置进行预测,加速了Mean Shift算法的收敛速度。本文剩余部分安排如下:第2节介绍Mean Shift跟踪算法,第3节介绍简化的LTP纹理模型,然后给出本文算法的实现和实验结果分析,最后引出结论。

2 Mean Shift目标跟踪

Mean Shift算法的基本思想就是不断地把跟踪窗口中心向样本均值移动使之收敛(重合)的过程。设 h_x , h_y 分别是跟踪窗口宽度和高度的一半, x_c 为模板的中心位置, x_i 为第 i 个像素点的位置, n 为目标区域像素的总个数,则目标的特征向量 \mathbf{u} 的概率估计为

$$\hat{q}_u = C_q \sum_{i=1}^n k(x_i - x_c, h_x, h_y) \delta[b(x_i) - \mathbf{u}] \quad (1)$$

类似地,以 y 为候选窗口中心的候选目标的特征向量 \mathbf{u} 的概率为

$$\hat{p}_u(y) = C_p \sum_{i=1}^n k(x_i - y, h_x, h_y) \delta[b(x_i) - \mathbf{u}] \quad (2)$$

其中 C_q 和 C_p 为归一化常数

$$C_q = 1 / \sum_{i=1}^n k(x_i - x_c, h_x, h_y) \quad (3)$$

C_p 与 C_q 的表达式相类似。 $b(x_i)$ 为目标的特征函数(索引函数)在像素点 x_i 的值, δ 函数的作用是判断 x_i 的索引值是否为特征值 u , $k(x)$ 为核函数。

在目标模板与候选区域进行匹配的过程中,通常采用下述Bhattacharyya系数来估计两区域密度

之间的相似程度:

$$\hat{\rho}(y) \equiv \rho[p(y), q] = \sum_{u=1}^m \sqrt{p(y)\hat{q}_u} \quad (4)$$

其中 m 为特征向量的总个数。两密度之间的分布越相似, $\hat{\rho}$ 值越大。则目标中心为^[10,11]

$$y = \frac{\sum_{i=1}^n x_i w_i k(y - x_i, h_x, h_y)}{\sum_{i=1}^n w_i k(y - x_i, h_x, h_y)} \quad (5)$$

其中 w_i 是加权系数

$$w_i = \sum_{u=1}^m \delta[b(x_i) - u] \sqrt{q_u / p_u(y)} \quad (6)$$

由于从式(5)得不到 y 的解析解,通常采用迭代方式求解。设初值为 y_0 ,则令 $y(0) = y_0$,利用式(5)进行迭代,直到 $\|y(k) - y(k-1)\|$ 小于某一阈值或达到最大迭代次数。对于进行较有规律运动且运动速度不高的目标,通过Mean Shift算法,就可以找到新的目标中心位置,实现对目标的跟踪。

3 LTP纹理模型

3.1 LBP纹理模型及其uniform模式

LBP是一种点样本纹理估计方法,点的纹理通常是通过该点的灰度值与其邻域点的灰度值做减法运算^[6,9]而得,对于坐标不是整数位置的采样点的灰度值,可以通过双线性插值法得到。图像的LBP纹理特征计算公式^[5,9,12]如下:

$$\text{LBP}_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p \quad (7)$$

其中 $s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$ 为阈值函数, R 是中心像素与其邻像素的距离,反映了纹理在空间上的分辨率。 P 为邻域像素的数目, g_c 表示中心点的灰度值, g_p 表示以 g_c 为中心点、以 R 为半径的圆环上的第 p 个分点的灰度值。本文选择 $P = 8$, $R = 1$ 。

式(7)只是具有尺度不变性,为了使纹理同时具有旋转不变性,可对上述二值模式按一定规律进行旋转,这样可把原来的256种纹理模式合并成36种。其中,出现频率最高的9种模式称为uniform模式(图1所示,黑点代表1,白点代表0),对应的模式值为0~8。记为 $\text{LBP}_{8,1}^{\text{uni}}$ (详见文献[12]),具体定义式如下:

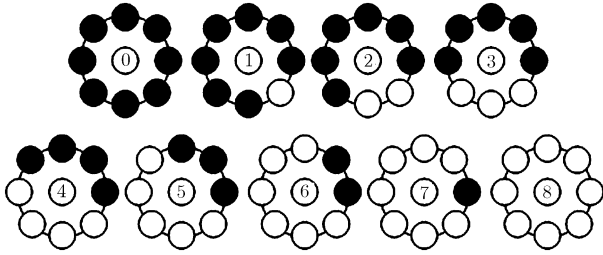
式(7)只是具有尺度不变性,为了使纹理同时具有旋转不变性,可对上述二值模式按一定规律进行旋转,这样可把原来的256种纹理模式合并成36种。其中,出现频率最高的9种模式称为uniform模式(图1所示,黑点代表1,白点代表0),对应的模式值为0~8。记为 $\text{LBP}_{8,1}^{\text{uni}}$ (详见文献[12]),具体定义式如下:

$$\text{LBP}_{8,1}^{\text{uni}}(x_c, y_c) = \begin{cases} \sum_{p=0}^7 s(g_p - g_c), & U(\text{LBP}_{8,1}) \leq 2 \\ 9, & \text{其他} \end{cases} \quad (8)$$

其中, $U(\text{LBP}_{8,1}) = |s(g_7 - g_c) - s(g_0 - g_c)| + \sum_{p=1}^7 |s(g_p - g_c) - s(g_{p-1} - g_c)|$ 为图1所示环中各点值0,1的变化次数。

3.2 LTP模型及其直方图的建立

上述的LBP模型虽然具有尺度不变性和旋转不变性,但对于噪声干扰较敏感,鲁棒性较差。为

图1 $LBP_{8,1}^{uni}$ 的 uniform 模式

了增加其抗噪声性能,文献[9]对 LBP 模型的阈值函数进行修改,提出了 LTP 纹理模式,其阈值函数 $s'(x_c, y_c)$ 定义如下:

$$s'(x_c, y_c) = \begin{cases} 1, & g_p \geq g_c + t \\ 0, & |g_p - g_c| < t \\ -1, & g_p \leq g_c - t \end{cases} \quad (9)$$

其中 t 为噪声门限, t 的大小直接决定了 LTP 算法的抗噪声性能。本文取 $t = 5$ 。LTP 纹理模式的对称性和噪声门限可以有效地滤除噪声,提高算法鲁棒性。然而, LTP 纹理模型的三值结构却大大增加了算法的复杂度。为了降低其算法复杂度,文献[9]把 LTP 纹理模式拆分成 ULBP(Upper LBP) 和 LLBP(Lower LBP) 两个 LBP 纹理模式进行处理。其中, ULBP 的阈值函数 $s_u(x_c, y_c)$ 为

$$s_u(x_c, y_c) = \begin{cases} 1, & s'(x_c, y_c) = 1 \\ 0, & \text{其他} \end{cases} \quad (10)$$

LLBP 的阈值函数 $s_l(x_c, y_c)$ 为

$$s_l(x_c, y_c) = \begin{cases} 1, & s'(x_c, y_c) = -1 \\ 0, & \text{其他} \end{cases} \quad (11)$$

其中 $s'(x_c, y_c)$ 是根据式(9)计算出的值。对这些纹理特征建立直方图,即可构成 LTP 纹理模型。

4 基于多特征目标模型的 Mean Shift 人脸跟踪算法

4.1 核函数的选择

考虑到越是外围的像素受背景或遮挡的影响的可能性越大,就越不可靠,本文采用如下矩形高斯核函数

$$k((\Delta x, \Delta y), h_x, h_y) = \begin{cases} \exp\{-[\Delta x^2/h_x^2 + \Delta y^2/h_y^2]/2\}/2\pi h_x h_y, & \Delta x \leq h_x, \Delta y \leq h_y \\ 0, & \text{其他} \end{cases} \quad (12)$$

其中 Δx , Δy 是该像素点到跟踪窗口中心的水平距离和垂直距离。该核函数具有单调递减的凸轮廓,

对目标中心处所加的权值最大,离目标中心越远,所加的权值越小。这种加权方法降低了周边像素的影响,增加了估计的鲁棒性。这时式(5)可化为

$$y = \frac{\sum_{i=1}^n x_i w_i \exp\{-[\Delta x^2/h_x^2 + \Delta y^2/h_y^2]/2\}}{\sum_{i=1}^n w_i \exp\{-[\Delta x^2/h_x^2 + \Delta y^2/h_y^2]/2\}} \quad (13)$$

4.2 卡尔曼滤波器目标预测

必须指出,上述的 Mean Shift 算法是在对 Bhattacharyya 系数一阶近似的基础上得出的,当目标运动速度较高时,如果采用上一帧目标的中心作为跟踪窗口的初值进行迭代,既要增加迭代时间,又可能丢失跟踪目标。为此,本文利用卡尔曼(Kalman)滤波器^[13]估计目标的中心位置作为迭代初值,既可减少目标丢失的可能性,又可以提高迭代收敛速度。必须指出,如果目标进行无规则运动可能导致卡尔曼预测失效。这时将以前一帧目标的位置作为跟踪窗口的初值,并扩大搜索范围对目标进行跟踪。如果仍然搜索不到目标,则重新启动人脸检测算法搜索目标。

4.3 多特征模型的建立

考虑到在 uniform 模式中,模式 0, 模式 1 表示噪声点,模式 7, 模式 8 表示暗点或光滑区域,模式 2, 模式 6 表示线段,模式 3, 模式 5 表示角点,模式 4 表示边界。其中, 0, 1, 7, 8 这 4 种模式所提供的信息量少,为次要模式;而其它的 5 种模式可用于表示图像的纹理信息,是关键模式。当把一个 LTP 纹理模式拆分成两个 LBP 模式时,所产生的模式冗余主要为模式 7 和模式 8。因此,本文把模式 0, 1, 7, 8 合并为一种模式,与其它的 5 种模式一起构成 6 个纹理特征,对 ULBP 和 LLBP 的关键模式建立直方图,就构成了一个简化的 LTP 纹理模型。该纹理模型既基本消除了从 LBP 升级到 LTP 产生的冗余,又增强了图像的关键纹理信息,克服了 LBP 纹理模式对噪声的敏感性,提高了纹理模型的准确性与鲁棒性。

颜色模型选用 HSV 颜色空间中的受光照影响不大的 H 分量作为目标的颜色特征向量,把 H 分量量化为 16 个色度空间。上述 LTP 直方图和 H 分量直方图的结合构成了本文的多特征模型。

4.4 算法描述

综上所述,本文提出的基于颜色和 LTP 纹理的多特征目标跟踪算法步骤如下:

(1)启动 AdaBoost 人脸检测算法,初始化跟踪目标和卡尔曼滤波器,并提取该目标的特征直方图模型,令 y_0 为该目标的中心位置。

(2)设定 Mean Shift 算法允许的最大迭代次数 N_0 (N_0 可通过实验确定,本文取 $N_0 = 10$)。

(3)以 y_0 为候选目标的中心位置,计算候选目标的特征直方图模型。

(4)根据式(6)计算候选目标区域内的各像素的权值。

(5)根据式(13)进行迭代计算目标的新位置 y 。若 $\|y - y_0\| < \varepsilon$ (本文取 $\varepsilon = 1.0$),则停止迭代,并令 $y_0 = y$,转步骤(7)。若没超过最大迭代次数,令 $y_0 = y$,转步骤(3)。

(6)若这轮迭代的初值采用的是卡尔曼滤波预测的初值,以前一帧目标的位置作为目标的初值,令最大迭代次数等于 $2N_0$,转步骤(3);否则认为目标丢失,转步骤(1)。

(7)对跟踪目标进行准确定位后,利用卡尔曼滤波器预测该目标在下一帧图像中的位置,即令 $y_0 = (x_p, y_p)$ (卡尔曼滤波器预测后目标的横坐标和纵坐标)。

(8)读取下一帧图像,转步骤(2)。

5 实验结果及分析

为了测试本文算法在动态场景下对目标跟踪的性能,本文选择多个 CIF 视频序列进行人脸跟踪测试,并给出对 Stefan 标准序列(共 90 帧)的实验结果(如图 2 所示)。实验中对帧图像未做任何预处理,

计算机配置为 P4 CPU 3.0 GHz,内存 1 GB,开发环境为 Visual C++。图 2 给出了 RGB 模型^[7](简称 RGB),文献[8]算法和本文算法对 Stefan 序列的跟踪结果。

从图 2 可以看出,对于背景比较复杂且运动速度比较快的 Stefan 序列,采用 RGB 方法,在第 50 帧时已经完全失去跟踪目标,跟踪失败,而文献[8]和本文算法都能跟踪到人脸目标。在对 Stefan 视频序列进行人脸跟踪时,文献[8]把 Stefan 序列中第 50 帧和第 89 帧的人脸定位到头部的左上角,这是由于 LBP 纹理模式对于中心点像素比较敏感,造成目标定位不准确。而本文算法解决了 LBP 纹理模型对于中心点像素比较敏感的问题,因此相对于前两种算法,本文算法具有更好的跟踪效果。

就计算复杂度而言,如果只是从单点特征向量的计算复杂度考虑,文献[8]和本文算法都需要考虑该点的邻域像素,而 RGB 模型只需将该点的颜色进行量化,加入相应特征空间即可。因此,文献[8]和本文算法的单点计算复杂度要高于 RGB 模型。但是文献[8]和本文算法不需要对目标的所有点都进行计算,只选取表示线段、角点和边界的模式点进行运算。例如,对于 Stefan CIF 序列中第 1 帧人脸目标



图 2 从上到下分别为:RGB,文献[8]和本文算法对 Stefan 序列中人脸的跟踪结果

来说, RGB 模型要对整个目标中的 357 个点全部进行计算, 但只有 203 个点属于本文算法中的模式点, 即本文算法只有 56.7% 的点参与计算。表 1 给出了 RGB, 文献[8]和本文算法对 Stefan 序列前 40 帧跟踪测试性能比较。从表 1 可以看出, 3 种算法的直方图总数分别为 4096, 320, 576 条。可见, 本文算法的直方图数与文献[8]的直方图数在同一数量级, 仅为 RGB 方法 14.1%。由于在 Mean Shift 算法中, 主要的计算量在于跟踪过程的迭代计算, 本文采用卡尔曼滤波器对目标进行预测, 大大减少了迭代次数, 提高了本文算法的跟踪速度。图 3 给出了上述 3 种算法对 Stefan 序列前 40 帧中目标正确跟踪的迭代次数曲线。实验结果表明, 本文算法相对于 RGB 模型和文献[8]在对目标的跟踪性能上有明显提高。

从上述实验结果可以看出, 本文提出的基于 LTP 纹理的多特征融合人脸跟踪算法具有良好的跟踪效果和鲁棒性, 即使在环境干扰比较大的情况下仍具有良好的跟踪效果, 能满足实时跟踪的要求。

表 1 3 种算法对 Stefan 序列前 40 帧中目标的跟踪性能比较

目标表示方法	直方图条数	平均每帧的迭代次数(次/帧)	平均跟踪速度(帧/s)
RGB 方法	16×16×16	4.625	35
文献[8]	8×8×5	3.025	42
本文方法	16×6×6	1.6	51

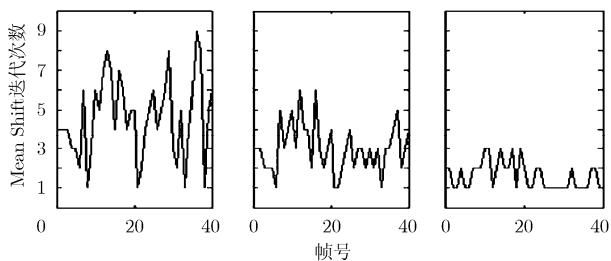


图 3 从左到右分别为: RGB, 文献[8]和本文算法的 Mean Shift 迭代次数示意图

6 结论

本文提出了一种基于纹理与颜色的多特征 Mean shift 人脸跟踪算法。首次把 LTP 纹理模式引入人脸跟踪, 有效地解决了人脸跟踪算法的鲁棒性问题。通过把 LTP 模式分解为两个 LBP 模式的方法降低了 LTP 纹理模式的计算复杂度。通过选取关键纹理模式, 合并次要纹理模式, 提出了简化 LTP 纹理直方图, 有效地去除了冗余, 增强了关键纹理模式, 进一步降低了计算复杂度。通过结合 LTP 纹理特征与颜色特征建立多特征人脸模型对 Mean Shift 人脸跟踪算法进行改进, 达到了良好的跟踪性能。实验验证了本文算法的正确性。

参考文献

- [1] Comaniciu D, Ramesh V, and Meer P. Real time tracking of non-rigid objects using mean shift[C]. Computer Vision and Pattern Recognition, Hilton Head Island, SC, USA, Jun. 13-15, 2000, Vol. 2: 142-149.
- [2] Wu Y and Huang T S. Robust visual tracking by integrating multiple cues based on co-inference learning[J]. *International Journal of Computer Vision*, 2004, 58(1): 55-71.
- [3] Triesch J and Von der Malsburg C. Self-organized integration of adaptive visual cues for face tracking[C]. Automatic Face and Gesture Recognition, Grenoble, France, Mar. 28-30, 2000: 102-107.
- [4] Xu X and Li B. Head tracking using particle filter with intensity gradient and color histogram[C]. Conference on Multimedia and Expo, Amsterdam, Netherlands, Jul. 6-9, 2005: 888-891.
- [5] Ojala T, Pietikainen M, and Harwood D. A comparative study of texture measures with classification based on feature distribution[J]. *Pattern Recognition*, 1996, 29(1): 51-59.
- [6] Nguyen Q A, Antonio R K, and Shen C H. Enhanced kernel-based tracking for monochromatic and thermographic video[C]. Advanced Video and Signal Based Surveillance, Sydney, Australia, Nov. 11, 2006: 28.
- [7] 王永忠, 梁彦, 赵春晖, 等. 基于多特征自适应融合的核跟踪方法[J]. 自动化学报, 2008, 34(1): 393-399.
- [8] Wang Y Z, Liang Y, and Zhao C H, et al. Kernel-based tracking based on adaptive fusion of multiple cues[J]. *Acta Automatica Sinica*, 2008, 34(4): 393-399.
- [9] 宁纪锋, 吴成柯. 一种基于纹理模型的目标跟踪算法[J]. 模式识别与人工智能, 2007, 20(5): 612-618.
- [10] Ning J F and Wu C K. A mean shift tracking algorithm based on texture model[J]. *Pattern Recognition and Artificial Intelligence*, 2007, 20(5): 612-618.
- [11] Tan X Y and Bill T. Enhanced local texture feature sets for face recognition under difficult lighting conditions[C]. Analysis and Modeling of Faces and Gestures, Rio de Janeiro, Brazil, Oct. 20, 2007: 168-182.
- [12] Comaniciu D, Ramesh V, and Meer P. Kernel-based object tracking[J]. *Pattern Analysis and Machine Intelligence*, 2003, 25(5): 564-577.
- [13] Jiang Z L, Li S F, and Gao D F. An adaptive mean shift tracking method using multiscale image[C]. Wavelet Analysis and Pattern Recognition, Beijing, China, Nov. 2-4, 2007, Vol. 3: 1060-1066.
- [14] Ojala T, Pietikainen M, and Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. *Pattern Analysis and Machine Intelligence*, 2002, 24(7): 971-987.
- [15] Dellaert F and Thorpe C. Robust car tracking using kalman filtering and bayesian templates[C]. Conference on Intelligent Transportation Systems, Pittsburgh PA, Etats-Unis, Oct. 15-17, 1997, Vol. 3207: 72-83.

张涛: 男, 1982年生, 硕士生, 研究方向为图像处理与模式识别。
蔡灿辉: 男, 1954年生, 教授, 博士, 主要研究领域为图像处理、模式识别和视频通信。