

# 基于颜色和变形模板的实时人体检测

王长军 朱善安

(浙江大学电气工程学院, 杭州 310027)

**摘要** 针对目前人体检测算法存在不能检测多角度人体目标以及实时性差等问题, 提出了一种在静态图像中实时检测任意角度人体目标的算法。该算法分别利用目标颜色和轮廓两类特征构造两种检测器。颜色检测器首先进行基于面部肤色和头部发色的彩色分割, 然后引入积分图像算法快速提取分割后的图像头部目标区域。轮廓检测器利用头肩轮廓形状的稳健性, 用参数化变形模板对头肩轮廓建模, 该模板由两个存在几何尺度和位置约束的椭圆构成, 再定义两个不同计算复杂度的模板匹配策略对人体头肩部分进行分级检测。最后利用上述两种检测器构建一个级联检测系统, 级联检测结构大大提高了算法的速度, 使算法可以对分辨率为  $352 \times 288$  的图像做 30fps 的实时检测, 实验结果表明, 该算法是切实有效的。

**关键词** 人体检测 彩色分割 头肩轮廓 积分图像 变形模板 多级检测

**中图分类号**: TP391.41 **文献标识码**: A **文章编号**: 1006-8961(2006)06-0861-06

## Color and Deformable Templates Based Real-time Pedestrian Detection

WANG Chang-jun, ZHU Shan-an

(College of Electrical Engineering Zhejiang University, Hangzhou 310027)

**Abstract** Most of the present pedestrian detecting algorithms can not deal with multi-view objects and can not perform in real time. Faced with these problems, this paper provided a real-time multi-view pedestrian detecting algorithm, which utilizes respectively the color and the contour features to construct two detectors. One is a color based detector, which, first, segments the original images based on skin color of face and hair color, and then locates the candidate head regions by using integral images, and the other is the contour detector, which employs the robust head-shoulder contour feature to construct a deformable template based on two ellipses with size and position constraints. Two template matching algorithms with different computing complexity were introduced to detect head-shoulder objects. Finally, a cascaded detecting system was generated by the two detectors mentioned above. The whole system can process images with  $352 \times 288$  resolution at a speed of 30fps by using the cascaded structure. The algorithm proved to be effective.

**Keywords** pedestrian detection, color segmentation, head-shoulder contour, integral image, deformable template, cascaded detection

## 1 引言

人体 (pedestrians) 检测就是根据人体的某些特征在静态图像中检测人体目标, 实时人体检测算法具有广泛的应用领域, 如视频监控<sup>[1]</sup>、智能交通、辅助驾驶<sup>[2]</sup>、图像检索等。目前大多数人体跟踪算

法<sup>[3,4]</sup>都需要已知目标的初始位置(如通过人机交互方式), 而实用的跟踪系统需要自动定位初始目标, 这就需要人体检测算法的支持。检测人体目标比检测其他目标更具挑战性。第一, 人体目标是典型的非刚性目标, 可能呈现出各种不同形状, 单一的模型几乎不可能捕获各种形状的人体; 第二, 人的不同衣着, 使基于区域特征的算法也不能有效地工作。

收稿日期: 2005-05-16; 改回日期: 2005-07-25

第一作者简介: 王长军(1979 ~ ), 男, 2001年于浙江大学获学士学位, 现为浙江大学电气工程学院博士研究生。主要研究方向为模式识别、智能监视等。E-mail: wangzju@126.com

Mohan 等人<sup>[5]</sup>训练 4 个 SVM (支持向量机) 分类器分别用来检测特征相对稳健的头、双腿、左臂和右臂, 然后通过几何约束得到人体目标。该方法由于引入多个 SVM 分类器, 使得训练分类器的工作量和检测时计算量非常大, 难以在 PC 处理器或 DSP (digital signal processor) 上实现实时检测, 而且该方法也只限于正面和背面人体, 对于侧面人体无能为力。Sidenbladh 以人在行走时的光流为特征训练一个 SVM 分类器<sup>[6]</sup>, 由于引入光流特征, 解决了由于不同衣着而产生的区域特征不稳健问题。但该方法只适用于背景静止的应用场合, 对于如辅助驾驶等背景运动的场合不再有效。还有一些算法用人脸检测代替人体检测, 人脸相对于人体而言, 具有形状变化小, 区域特征稳健的优点, 但人脸检测要求图像的解析度高, 且只能是正面人脸, 或小于 90° 的侧面人脸<sup>[7]</sup>, 对于大于 90° 甚至是背面的人体目标, 便失去作用。

本文提出一种基于颜色分割和变形模板的可以在静止图像中实时检测各个角度人体目标的算法。该算法分别利用头部颜色和头肩轮廓特征构造两种检测器。第 1 种为基于颜色检测器, 它首先在  $YC_b$  色彩空间对头部颜色特征建模, 利用该模型对原图像做彩色分割, 然后引入积分图像算法快速提取头部颜色特征的区域; 第 2 种为基于变形模板的轮廓检测器, 它充分利用人体头肩轮廓特征相对稳健的特点, 将头肩轮廓建模为由两个具有几何约束的椭圆构成的参数化变形模板, 该模板具有参数少, 建模准确, 计算量小等优点。通过极少的参数变化就可以表达多角度头肩轮廓形状, 并定义二种不同计算量的模板匹配算法, 分别对目标做由粗到精的匹配, 最后将上述两种检测器组成级联检测系统, 该系统输出为人体目标头肩位置。由于采用了一系列的方法加快算法检测速度, 算法对于 352 × 288 分辨率的图像可以达到 30fps 的检测速度, 具有实用价值。

## 2 颜色检测器

不同的衣着造成人体目标的颜色特征存在极大的不稳健性, 但人的头部颜色相对稳定, 基本由面部肤色和头部发色组成。颜色检测器即对肤色和发色进行建模, 然后对原始图像做彩色分割, 最后引入积分图像算法快速提取头部区域。

### 2.1 颜色检测

对肤色进行建模需要选取一个合适的色彩空

间, 使肤色在这个空间中有很好的聚类性。较常用的方法是在归一化 rgb 空间对肤色聚类建立 2D 的 Gaussian 概率密度模型<sup>[8]</sup>, 但这种方法效果并不理想, 原因是肤色的色差分量并非完全独立于亮度分量, 而是非线性地依赖于亮度分量<sup>[9]</sup>, 本文算法采用  $YC_b$  空间作为肤色检测空间, 此空间与人眼对色彩的感知一致, 而且将亮度和色差分离开来, 对肤色有很好的聚类性能, 通过对  $YC_b$  空间进行非线性变换, 使肤色的聚类性独立于亮度, 从而获得理想的肤色检测效果。

色差  $C_r$  和  $C_b$  可分别看作亮度  $Y$  的函数  $C_r(Y)$  和  $C_b(Y)$ 。设变换后的色差分量为  $C'_r(Y)$  和  $C'_b(Y)$ 。变化公式如下<sup>[11]</sup>:

$$C'_i(Y) = \begin{cases} \text{若 } Y < K_l \text{ 或 } K_h < Y, \text{ 则} \\ (C_i(Y) - \bar{C}_i(Y)) \cdot \frac{W_{C_i}}{W_{C_i}(Y)} + \bar{C}_i(K_h) \\ \text{若 } Y \in [K_l, K_h], \text{ 则} \\ C_i(Y) \end{cases}$$

$$W_{C_i}(Y) = \begin{cases} WL_{C_i} + \frac{(Y - Y_{\max}) \cdot (W_{C_i} - WL_{C_i})}{K_l - Y_{\min}} & Y < K_l \\ WH_{C_i} + \frac{(Y_{\max} - Y) \cdot (W_{C_i} - WH_{C_i})}{Y_{\max} - K_h} & K_h < Y \end{cases}$$

$$\bar{C}_b(Y) = \begin{cases} 108 + \frac{(K_l - Y) \cdot (118 - 108)}{K_l - Y_{\min}} & Y < K_l \\ 108 + \frac{(Y - K_h) \cdot (118 - 108)}{Y_{\max} - K_h} & K_h < Y \end{cases}$$

$$\bar{C}_r(Y) = \begin{cases} 154 + \frac{(K_l - Y) \cdot (154 - 144)}{K_l - Y_{\min}} & Y < K_l \\ 154 + \frac{(Y - K_h) \cdot (154 - 132)}{Y_{\max} - K_h} & K_h < Y \end{cases}$$

式中,  $C_i$  表示  $C_r$  或  $C_b$ ;  $Y, C_r, C_b \in [0, 255]$ ,  $W_{C_r} = 38.76$ ,  $W_{C_b} = 46.97$ ,  $WL_{C_r} = 20$ ,  $WL_{C_b} = 23$ ,  $WH_{C_r} = 10$ ,  $WH_{C_b} = 14$ ,  $K_l = 125$ ,  $K_h = 188$ ,  $Y_{\max} = 235$ ,  $Y_{\min} = 16$ 。在变换后的  $C'_r C'_b$  空间中, 肤色点聚类成下式表示的椭圆区域。

$$\frac{(x - ec_x)^2}{a^2} + \frac{(y - ec_y)^2}{b^2} = 1$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} C'_b - c_x \\ C'_r - c_y \end{bmatrix}$$

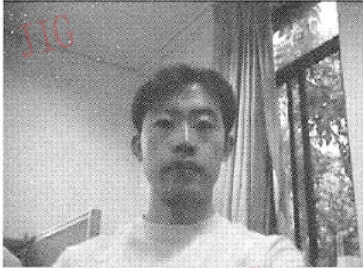
其中,  $c_x = 109.38$ ,  $c_y = 152.02$ ,  $\theta = 253$ ,  $ec_x = 1.60$ ,  $ec_y = 2.41$ ,  $a = 25.39$ ,  $b = 14.03$ 。

## 2.2 发色检测

头发颜色一般接近黑色,具有亮度分量  $Y$  较小,  $C_b$ ,  $C_r$  分量接近于零的特点,通过采样分析,发色点  $P_i$  聚类于以下区域:

$$\{P_i | 120 < Y(P_i) < 80, C_r(P_i) < 10, C_b(P_i) < 10\}$$

以上检测准则有可能把图像中灰度值较低的区域误检为头部区域,误检区域绝大部分会在后续的轮廓检测器中去除。图 1(b)为对图 1(a)做肤色和发色联合检测效果。



(a) 原始图像



(b) 检测结果

图 1 肤色发色检测结果

Fig. 1 Skin and hair tone detection

## 2.3 基于积分图像的目标定位

肤色和发色检测器将原始图像分割为由目标和非目标像素组成的二值图像,下一步要对此二值图像中目标做定位。有几种算法可以用来定位目标,如 K 均值算法、连通性分析算法等,但都存在计算量大的缺点。本文算法采用简单计算矩形子窗口内目标像素数量的方法判断该子窗口中是否包含目标。若用常规求和算法,有很大的计算量。本文引入积分图像,它可以大大减轻计算负荷。积分图像算法由 Paul 等人在基于 HARR 特征的目标检测算法中提出<sup>[10]</sup>。

积分图像中  $(x, y)$  点像素值为原图像左侧和上

方全部像素值的和,即

$$I'(x, y) = \sum_{x' \leq x, y' \leq y} I(x', y')$$

式中,  $I'$  为积分图像,  $I$  为原图像。可以通过下式迭代计算得到积分图像。

$$s(x, y) = s(x, y - 1) + I(x, y) \quad (1)$$

$$I'(x, y) = I'(x - 1, y) + s(x, y)$$

得到积分图像后,原图像中子窗口内像素值之和  $sum$  可以通过下式计算:

$$sum = I'(x_4, y_4) + I'(x_1, y_1) - I'(x_2, y_2) - I'(x_3, y_3)$$

式中,  $(x_1, y_1)$ ,  $(x_2, y_2)$ ,  $(x_3, y_3)$ ,  $(x_4, y_4)$  分别为子窗口左上、右上、左下、右下顶点坐标,因为对于一幅目标图像,积分图像只需计算一次,因此大大减小了计算量。

## 3 轮廓检测器

目标识别的特征一般分为两类:区域特征(如颜色、纹理等)和轮廓特征。人体不同衣着导致区域特征不稳健,相对而言,轮廓形状特征更稳健,注意到虽然人体目标呈现强烈的非刚性特点,但人体肩部及肩部以上区域的轮廓形状基本稳定,只是在不同侧面肩部形状会有较大变化,因此本文算法采用参数化变形模板<sup>[11]</sup>对头部轮廓进行建模,然后采用多级的模板匹配算法进行目标检测。本算法与 Birchfield 采用的算法<sup>[12]</sup>有相似之处,不同的是 Birchfield 只做头部跟踪,而且不具备检测能力,头部目标的识别与初始定位通过人机交互方式实现,这就降低了对算法的要求,只需简单地对头部轮廓建模,而且不必对整个图像做模板匹配,这就减小了计算量,而本算法设计用于在静态图像中实时检测人体目标,因此首先要实现头部目标的识别,其次要算法具有较高的检测速度。

### 3.1 构造变形模板

变形模板大体可以分为参数化和非参数化变形模板。非参数化变形模板不需要目标形状的先验知识,比较典型的是活动轮廓模型<sup>[13]</sup>。参数化模板,通过先验知识将目标轮廓可能的形状表达为由几个参数确定的一簇曲线。目标检测就是模板匹配过程,它通过最小化某个能量函数而达到。对于参数化变形模板,模板的自由度即参数的个数直接影响算法的性能。自由度太多会导致匹配计算量增大,影响实时性,自由度太少,又难以准确建模目标形状。本文算法用两簇具有几何尺度和位置约束的椭

圆弧段对头肩轮廓建模。该模板可以准确表达头肩轮廓的各种变化,同时具有很小的计算量。用  $e(x, y, \gamma, a, \alpha, \beta)$  表示椭圆模型(如图 2),其中,  $e(\cdot)$  表示椭圆簇,  $(x, y)$  为椭圆圆心坐标,  $\gamma$  为椭圆垂直与水平轴之比,  $a$  为表示垂直轴长,  $\alpha$  为弧段起始角,  $\beta$  为弧段终止角。

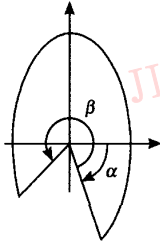


图 2 椭圆模板

Fig. 2 Ellipse template

头部模型为  $e_1(x_1, y_1, \gamma_1, a_1, \alpha_1, \beta_1)$ , 肩部模型为  $e_2(x_2, y_2, \gamma_2, a_2, \alpha_2, \beta_2)$ 。实验采用 Sobel 算子提取目标边缘,结果表明,接近颈部的边缘较弱,因此本文算法去除头部模板  $e_1$  下部部分弧段,取  $\alpha_1 = -75, \beta_1 = 225$ , 同样,肩部模板  $e_2$  的下半部分不能表达目标轮廓,因此只保留  $e_2$  上半部,取  $\alpha_2 = 0, \beta_2 = 180$ 。对于不同的人,头部形状  $\gamma_1$  变化不会太大,为减少模板匹配的计算复杂度,  $\gamma_1 = 1.2$ 。

如图 3 所示,  $e_1$  和  $e_2$  之间的几何位置约束满足

$$\begin{cases} a_2 = d - s \\ b_2 = a_2 / \gamma_2 \end{cases} \quad (2)$$

尺度约束满足

$$\begin{cases} y_2 = y_1 + d + \delta \\ x_2 = x_1 + \varepsilon \end{cases} \quad (3)$$

式中,  $d = ka_1, s = ha_1$ 。  $a_1, \delta, \varepsilon, \gamma_2$  构成变形模板参数空间。图 4 分别显示了由参数  $\varepsilon$  和  $\gamma_2$  变化得到的变形模板,  $\varepsilon$  表达人体侧面投影情况下,头部和肩部中心发生的位移,  $\gamma_2$  表达人体由正面或背面转到侧面过程中,肩部形状发生的变化。

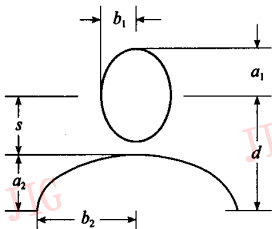


图 3 头肩模板几何关系

Fig. 3 Head-shoulder position

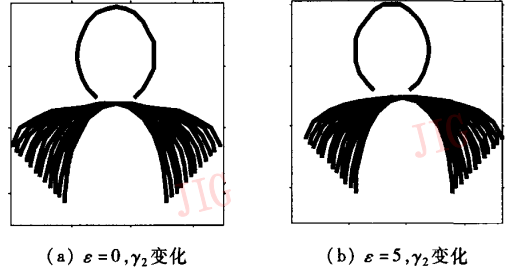


图 4 头肩变形模板

Fig. 4 Head-shoulder template

### 3.2 模板匹配

基于椭圆的变形模板有效地降低了人体这一非刚性目标的表示复杂度,减小了目标检测的计算量,但依然具有相当的计算量,因此引入两种匹配算法,分别对目标进行由粗到精的检测,再次减小匹配计算量,提高检测速度。粗匹配算法采用模板中坐标点处图像梯度幅值的均值衡量匹配程度,定义能量函数

$$\phi(x_i, y_i) = \frac{1}{N_e} \sum_{i=1}^{N_e} |\nabla I(x_i, y_i)| \quad (x_i, y_i) \in e_m \quad (4)$$

式中,  $I(x, y)$  为像素  $(x, y)$  的灰度,  $i$  为椭圆模板中点的索引,  $N_e$  为组成模板的点数,  $e_m$  表示椭圆模板  $e_1$  或  $e_2$ 。精匹配算法将梯度方向信息引入能量函数,定义

$$\phi_g(x_i, y_i) = \frac{1}{N_e} \sum_{i=1}^{N_e} |n_e(x_i, y_i) \cdot \nabla I(x_i, y_i)|$$

式中,  $n_e(x, y)$  为模板在点  $(x, y)$  的单位法向量。

首先让图像通过  $\phi$  检测器,计算量小,可以快速检测出候选目标,由于只考虑梯度幅值信息,对噪声的鲁棒性不强,将候选目标通过  $\phi_g$  进行精匹配,由于  $\phi_g$  既考虑梯度幅值又考虑梯度方向信息,因此有很强的噪声抑制能力。

### 4 完整检测器结构

目前多数的快速检测系统均由多个检测级联构成<sup>[10]</sup>,每个检测器被设计用来检测目标的某一特征,算法简单,计算量小,检测速度快,每一级检测器都可以去除一定数量的非目标区域,然后把通过检测的区域送给下一级检测器,因此在不降低检测性能的情况下提高了整个系统的检测速度。

本检测系统也采用级联检测结构,由一个肤色检测器,两个头模板检测器和两个肩模板检测器级

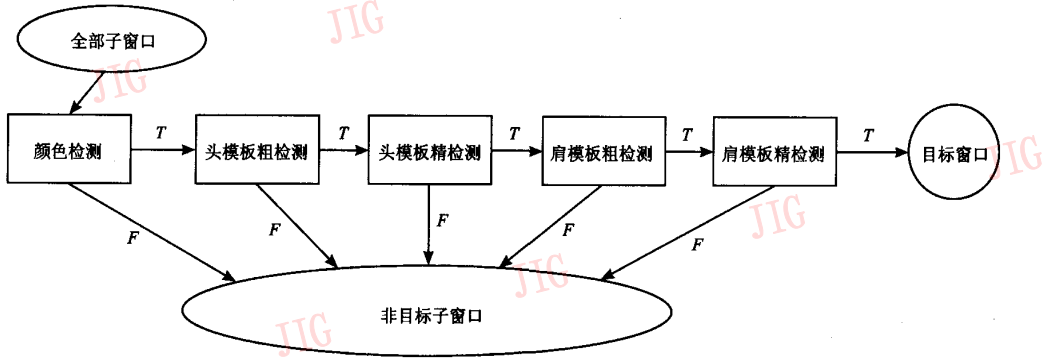


图 5 级联检测系统

Fig.5 Cascaded detection system

联而成。如图 5 所示,系统输入为图像中某一范围内尺度的全部子窗口,将子窗口依次通过颜色检测,头检测和肩检测,能量函数值若超过某一阈值则进入下一级检测器,否则被丢弃,最后输出目标位置。

对不同侧面呈现出的宽度,实验中取  $\gamma_2 \in [0.4, 0.6]$ ; 常数  $k$  决定头和肩椭圆圆心相对位置,实验中取  $k = 2.5$ ,  $h$  决定人的颈部长度,实验中取  $h = 1.2$ 。

为提高算法速度,采用长短步结合的方法枚举子窗口,长步跨度为 3 个像素,在能量函数值较低时采用,可以快速扫描大面积的非目标区域,当能量函数达到某个阈值后,切换短步枚举子窗口,以准确定位目标,短步跨度为 1 个像素。

采用连接于 PC 机上的 CCD 摄像机实时采集办公室环境下的视频信息对算法的有效性进行测试。图 6 展示了几幅较有代表性的检测情形,实验结果

### 5 实验结果与讨论

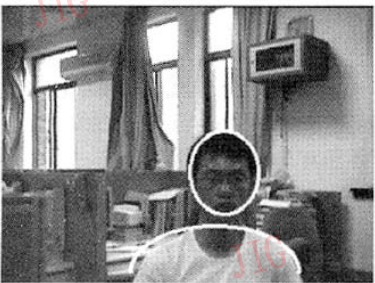
变形模板参数  $a_1$  决定要检测人体目标投影在图像平面中的尺度,单位为像素,实验中取  $a_1 \in [10, 20]$ ; 参数  $\delta$  和  $\varepsilon$  分别决定肩部相对于头部的纵向和横向偏移量,实验中取  $\delta \in [-4, 4]$ ,  $\varepsilon \in [-4, 4]$ ;  $\gamma_2$  决定肩部



(a) 45°侧面情形



(b) 背面情形



(c) 正面情形



(d) 90°侧面情形

图 6 人体检测结果

Fig.6 Experiment results

表明,算法可以对视频中的各角度人体目标进行正确的检测,检测速度平均可达 30fps。需要指出的是,本算法在非常嘈杂的背景下也存在一定的误检测。

## 6 结 论

针对目前人体检测算法计算量大,实时性差,难以处理多角度人体目标的问题,提出了一种基于颜色和轮廓形状特征的人体目标实时检测算法。该算法首先利用头部的肤色和发色构造  $YC_bC_r$  空间的头部检测器,其中,肤色检测算法利用非线性变换使色差分量独立于亮度分量,从而使用肤色检测效果更好。然后利用头肩轮廓形状相对稳定的特点,构造基于椭圆变形模板的头肩轮廓检测器,通过不同的参数取值,该模板可以准确地表达不同角度的人体目标轮廓,使算法具备了检测多角度人体目标的能力。算法最后用级联检测结构组成检测系统,实现目标的快速检测。

采用一系列方法提高检测算法的速度,包括积分图像算法、简单有效的椭圆变形模板、由粗到精的模板匹配算法、级联检测结构及长短步结合的子窗口枚举策略等,使检测算法的速度达到 30fps。

本文算法同其他检测一样存在一定的误检率,由于该算法的快速性,可以考虑增加前处理或后处理算法以降低误检率。另外,借助该算法运行快速的优势,可将该算法用于人体运动的跟踪<sup>[14]</sup>,在这方面已做了成功的尝试。

## 参考文献 (References)

- Elegmmal A, Duraiswami R, Harwood D, *et al.* Background and foreground modeling using nonparametric kernel density estimation for visual surveillance [J]. *Proceedings of the IEEE*, 2002, **90** (7): 1151 ~ 1163.
- Mckenna S J, Jabri S, Duric Z, *et al.* Tracking interacting people [A]. In: *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition [C]*, Grenoble, France, 2000: 348 ~ 353.
- Paragios N K, Deriche R. A-PDE-based level-set approach for detection and tracking of moving objects [A]. In: *Proceedings of Sixth International Conference on Computer Vision [C]*, Washington, DC, USA, 1998: 1139 ~ 1145.
- Comaniciu D, Ramesh V, Meer P. Real-time tracking of non-rigid objects using mean shift [A]. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C]*, South Carolina, USA, 2000: 142 ~ 149.
- Mohan A, Papageorgiou C, Poggio T. Example-based object detection in images by components [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, **23** (4): 349 ~ 361.
- Sidenbladh H. Detecting human motion with support vector machines [A]. In: *Proceedings of the 17th International Conference on Pattern Recognition [C]*, Cambridge, MA, UK, 2004: 188 ~ 191.
- Zhang Zhen-qi, Zhu Long, Li S Z, *et al.* Real-time multi-view face detection [A]. In: *Proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition [C]*, Washington, DC, USA, 2002: 142 ~ 147.
- Chong L, Tat-Seng C, Teck-Khim N. Face tracking in video with hybrid of Lucas-Kanade and condensation algorithm [A]. In: *Proceedings of 2003 International Conference on Multimedia and Expo [C]*, Maryland, USA, 2003: 293 ~ 296.
- Rein-Lien Hsu, Abdel-Mottaleb M, Jain A K. Face detection in color images [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24** (5): 696 ~ 706.
- Viola P, Michael J. Robust Real Time Object Detection [R]. Technical Report: CRL-2001-1, Cambridge University, Cambridge, MA, UK, 2001.
- Jain A K, Yu Z, Lakshmanan S. Object matching using deformable template [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1996, **18** (3): 267 ~ 278.
- Birchfield S. Elliptical head tracking using intensity gradients and color histograms [A]. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition [C]*, California, USA, 1998: 232 ~ 237.
- Kass M, Witkin A, Terzopoulos D. Snake: Active contour models [A]. In: *Proceedings of the 1st International Conference on Computer Vision [C]*, London, UK, 1987: 259 ~ 268.
- Shakhnarovich G, Viola P, Moghaddam B. A unified learning framework of real time face detection and classification [A]. In: *Proceedings of IEEE Conference on AFG [C]*, Washington DC, USA, 2002: 14 ~ 21.