

基于组合模板的人体头部精确定位算法

汤金宽¹⁾ 曹丹华¹⁾ 吴裕斌¹⁾ 顾雯雯²⁾

¹⁾(华中科技大学光电子科学与工程学院, 武汉 430074) ²⁾(西北工业大学理学院, 西安 710072)

摘要 对人体头部的准确定位是当今计算机视觉研究中一个非常重要的领域。在基于视觉特征的人体头部提取算法的基础上,提出了一种改进的算法——“粗—精”双步头部提取算法。该算法首先利用人体头部与躯干相对关系初步确定人体头部位置及大小,然后通过水平和垂直方向的两个精确定位模板对粗定位结果进行调整,最终达到精确定位的目的。详细阐述了“粗—精”双步头部提取算法的基本原理和实现方法,并针对行走中的人体,给出了改进前后两种算法实验效果。实验结果表明,改进后的头部定位算法可以在复杂情况下更精确地定位头部,鲁棒性强。

关键词 人体头部提取 模板匹配 精确定位 实时性 积分图

中图分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2007)08-1389-06

A Novel Head-location Algorithm Based on Combined Mask

TANG Jin-kuan¹⁾, CAO Dan-hua¹⁾, WU Yu-bin¹⁾, GU Wen-wen²⁾

¹⁾(Department of Optoelectronic Engineering, Huazhong University of Science and Technology, Wuhan 430074)

²⁾(School of Natural and Applied Science, Northwestern Polytechnical University, Xi'an 710072)

Abstract Human head location plays an important role in applications such as video surveillance, human activities analysis, perceptual user interface and so on. In this paper, A novel algorithm, called “coarse-accurate” two-step head location, based on view-based human head location algorithm, is proposed. The proposed approach first locates the head's coarse position depending on the relative relationship of the human head and torso, then adjusts the coarse position to the accurate position by adopting two novel masks. In the last part of this paper, the results of improved and unimproved algorithms are compared. The experimental results show that the proposed algorithm is more robust in complex conditions and can detect the head more accurately.

Keywords human head detection, mask matching, accurately location, real-time, integral image

1 引言

对人体头部大小以及位置的自动提取是当今计算机视觉研究中一个非常重要的内容。其被广泛的应用于很多涉及图像理解的领域,例如:视频监控系統、人体行为识别系統、虚拟人机界面、人体脸部自动识别等等。人体头部的精确定位,可以为更高层的图像理解算法例如脸部识别、行为分析等等提供输入,它提供的结果准确与否将直接影响着后续算法的成功率以及精确度。因此近几年来,头部提取

算法越来越受到了研究工作者的重视。

目前比较有效的头部提取算法主要可以分成二类:第一类是基于面部肤色信息的^[1-3],这类算法主要是利用通常场景中出现的目标人体面部一般是裸露的,因此充分利用人体肤色这一特征来定位头部。它的主要问题是需要大量的先验数据来统计得出人体肤色的分布,同时由于不同肤色人种、光照等等因素的影响常常会使得算法失败。第二类是基于头部外形信息的,这类算法一般是通过头部与身体的相对位置以及颈部特征来定位头部的。例如文献[4]利用头部与人体总的高度之比来完成定位,它要求人体

收稿日期:2005-10-10; 改回日期:2006-05-16

第一作者简介:汤金宽(1981~),男,华中科技大学光电子科学与工程学院物理电子学硕士研究生。主要研究工作是数字图像处理。
E-mail: tomgw@ hotmail. com

全部出现场景中同时其受阴影较大;文献[5]是通过找人体外轮廓的最小负曲率点定位出颈部,从而完成头部提取,它要求目标的外轮廓线比较光滑同时颈部特征明显;文献[6]是利用头部与躯干形成一个“凸”字形的特点来定位头部,这种算法实现效率高,同时由于其没有利用肤色、颈部、轮廓等等特征,因此其定位不易受到外界因素的干扰,成功率高同时比较准确。但是在实验中发现,此算法由于假定了“凸”字形的结构,当人体完成转身等动作时头部偏离躯干的中线,就很容易导致头部定位不准确甚至失败。因此对头部与躯干部分的相对位置要求较高。

本文在文献[6]的基础上,提出了一种新型的“粗一精”双步头部提取算法。它改进了原算法的不足之处,采用基于组合模板的双步定位法取代单模板的单步定位算法。实验结果表明,本文提出的算法不仅可以在正常情况下准确地定位头部,同时当由于人体行走、转身等原因导致头部偏离躯干中线,或者躯干部分提取不完整时,依然可以准确地完成提取,大大提高了算法的适应性。

2 文献[6]头部定位算法原理及存在问题

2.1 文献[6]头部定位算法原理

文献[6]头部定位算法基于视觉特征,将头部和躯干的相对大小和位置关系看成一个“凸”形结构,故构建的“凸”型模板如图 1 所示。

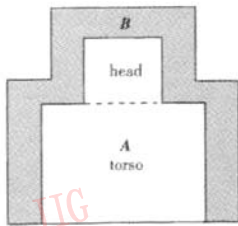


图 1 文献[6]算法模板

Fig. 1 The mask of Antonio Micilotta algorithm

可以看到此模板分成了两个区域:A 区域和 B 区域,其中,A 区域为真正人体头部与躯干的位置,而 B 区域则为模板的外部限制区域。

为此模板构造一个评价函数:

如果 $S \leq 0$, 则

$$S = \frac{1}{S(A)} (\sum A - \sum B) \quad (1)$$

其中,

$$S(A) = L_{\text{torso}}^2 + L_{\text{head}}^2 \quad (2)$$

$$\sum A = \sum_{(x,y) \in A} I(x,y) \quad (3)$$

$$\sum B = \sum_{(x,y) \in B} I(x,y) \quad (4)$$

而 $I(x,y)$ 是前景的二值化图像,其定义为

$$I(x,y) = \begin{cases} 1 & (x,y) \text{ 为前景区域} \\ 0 & (x,y) \text{ 为背景区域} \end{cases}$$

L_{torso} 和 L_{head} 分别是模板中头部区域和躯干区域的长度,因此 $S(A)$ 实际上即模板 A 区域的面积,此模板的一个特点是构造的 A 区域和 B 区域有: $S(A) = S(B)$ (其中 $S(B)$ 是区域 B 的面积)。将此模板用于前景的二值化图像 $I(x,y)$, 分别计算得出 $\sum A$ 和 $\sum B$, 然后代入评价函数计算其函数值。改变模板在前景二值化图像中的位置,同时分别改变模板的头部以及躯干部分的大小 (L_{torso} 和 L_{head})。计算各种情况下模板的评价函数值,然后找到函数值最高的模板,那么此时模板的位置以及大小就可以提取人体的头部的位置以及大小。

由评价函数的表达式可以看出,其函数值的变化范围是 $[0,1]$ 。若要使得评价函数的值最高,则需要 $\sum A \approx S(A)$, 同时 $\sum B$ 尽可能的接近零。在匹配过程中,若模板的位置出现在前景区域以外,其函数值很明显将会为零或者很低;若模板过大,则虽然 $\sum A - \sum B$ 的值会比较大,但是由于 $S(A)$ 的归一化的作用,使得评价函数值依然比较小;若模板过小,则由于 $\sum A$ 和 $\sum B$ 的值会很接近,于是 $\sum A - \sum B$ 的值会很小,导致评价函数值亦会很小。只有当模板的位置以及大小都比较合适的时候(如图 2),此时满足 $\sum A \approx S(A)$ 和 $\sum B \approx 0$, 使得评价函数值最大(接近于 1)。采用这种模板匹配

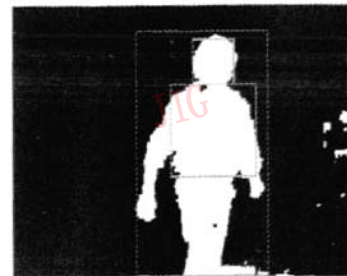


图 2 文献[6]头部定位效果图

Fig. 2 The result of Antonio Micilotta^[6]

的方法搜索出来的头部位置、大小以及躯干的位置在正常情况下是比较准确的。

2.2 算法的不足

引言部分已经提到,该模板匹配算法存在一个很大的弊端:其模板假定了头部位于躯干的中线部位。也就是说,它限定了头部和躯干的相对位置关系。此限制导致了在某些复杂情况下,如侧面转身时头部偏离了躯干的中心,脖子较长时头部

与躯干分离,躯干提取不完整等,定位头部位置的不准确。

分析导致“凸”型模板匹配定位头部失败的主要情况之后,可以将这种定位失败的原因归纳如下:

- (1) 垂直方向头部与躯干中线的偏离(如图3(a));
- (2) 水平方向头部与躯干中线的偏离(如图3(b))。

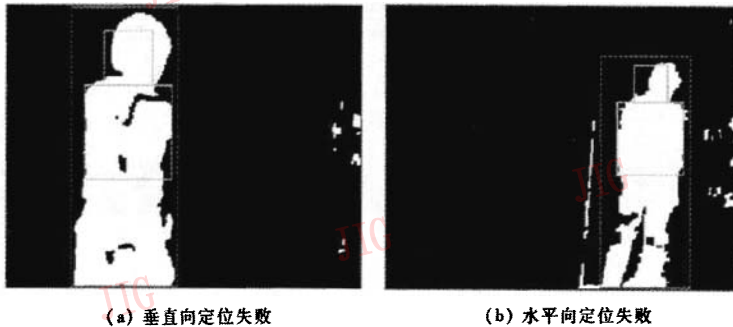


图3 头部定位失败效果

Fig. 3 The failed results of head location

这种定位误差对如视频监控、人体行为分析等应用的后续算法有着很大影响。例如对于视频监控系统在做人脸识别时,首先必须完成头部提取,如果头部提取不准确后续的算法就无法正常的运行。再如进行人体行为分析,需要建立人体模型,通常建立人体模型的算法都是从人体头部开始的,如果人体定位不准确,将对接下来定位其他肢体部分造成很大的障碍。本文基于模板定位的基本原理,针对运动人体识别,提出了改进的“粗一精”双步头部提取算法。

3 “粗一精”双步头部提取算法及实现

针对文献[6]算法的两个主要问题,在“凸”型模板初步定位出头部粗略位置后,另外构建了两个模板分别在水平方向和垂直方向小范围地搜索匹配,精确地定位出偏离躯干中线的头部位置。即先使用“凸”型模板完成粗定位,然后使用精确定位模板完成头部定位的调整。故称这种算法为“粗一精”双步头部提取算法。

3.1 前景提取算法

由于算法中模板匹配过程是基于二值化图像的,故首先必须完成人体前景区域的提取和二值化工作,本算法采用“背景图像差分法”完成前景

区域的提取与二值化。其算法原理是首先对背景场景用一种方法表现出来,然后将新的图像帧与该背景表现进行对比,从而侦测到场景中出现的异常运动。这种对背景场景的表现则称为“背景建模”。

通过分析发现,在理想情况下,相邻图像的同一背景区域的灰度值是一样的,其灰度值之差应该为零。但是由于噪声的存在,背景区域图像灰度实际上并不全为零。噪声因素包括由微小的相机抖动或数字化扫描引起的配准误差、照明及传感器的不稳定噪声等。通过分析一个静止场景(没有背景运动)中某一像素的灰度值在一段时间内(连续若干帧连续图像中)的变化规律,发现当图像的噪声变换规律可以用一个零均值的正态分布 $N(0, \sigma^2)$ 表示时,某一像素点灰度值的变换规律可以由一个均值为 μ 的正态分布 $N(\mu, \sigma^2)$ 来表示。而通过观察发现,大多数的噪声确实满足或者近视满足零均值的正态分布,所以本系统假设静止场景中,每一个像素点灰度值的变化规律均满足正态分布 $N(\mu, \sigma^2)$ 。因此可以根据正态分布的“ 3σ 原则”,利用背景场景中每个像素灰度值变化分布的均值 μ 以及方差 σ 进行背景建模。将每一帧新图像中的各个像素点与其对应的背景模型中像素的均值做差值,然后进行

阈值操作。为了获得完整的前景区域,还需进行影子的消除、中值滤波、形态学滤波、连通域分析等操作。

3.2 垂直向精确定位

垂直向精确定位的模板如图 4 所示。

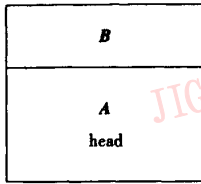


图 4 垂直向精确定位模板

Fig. 4 The vertical accurate location mask

若此时依然设定 B 区域面积与 A 区域面积相等,将会导致模板 B 区域面积过大。通过实验表明,当人体头部出现在靠近视场上方边缘时,这种情况将会导致模板的 A 区域无法与人体头部准确位置完全吻合,使得模板无法调整到垂直方向的最佳位置。因此为了使得模板不至于太大,选择 $S(B) = \frac{1}{4}S(A)$ 。

依然选择评价函数为

如果 $S_v \leq 0$, 则

$$S_v = \frac{1}{S(A)} \left(\sum A - \sum B \right) \quad (5)$$

这里 $\sum A$ 、 $\sum B$ 和 $S(A)$ 的定义与上文相同。使此模板在垂直向搜索,计算评价函数值。若搜索起始位置有 $\sum B = 0$, 则说明有可能头部定位过高,需要向下方滑动模板,其搜索的终止的条件是 $\sum B$ 大于某一阈值,此时表明模板已经过多地进入到头部区域;若搜索起始位置上有 $\sum B \neq 0$, 则说明有可能头部定位过低,需要向上方滑动模板直至 $\sum B = 0$ 。此时意味着模板已经离开了头部区域。在搜索过程中,评价函数值最大处即是头部在垂直向调整后的最佳位置。

3.3 水平向精确定位

水平方向精确定位的模板如图 5 所示。

同样为了避免模板面积过大,选择 $S(B)$ 是 $S(A)$ 的四分之一。此模板区域 B 的高度是区域 A 的一半,并且区域 B 位于区域 A 的中部,这样设计的目的是为了阻止下方颈部以及躯干的干扰。如果将区域 B 的高度设计成和区域 A 一致,那么在模板

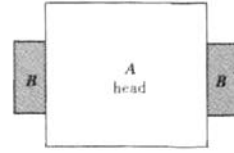


图 5 水平向精确定位模板

Fig. 5 The horizontal accurate location mask

水平向移动时,值 $\sum B$ 很容易受到躯干和颈部的影响,使得定位不再精确。同时由于人体的头部可以近视看成是一个长轴是垂直方向的椭圆,其水平宽度的最大值应该出现在头中部附近。基于以上两个理由,故将水平向精确定位模板设计成图 5 的形状。

依然设定评价函数为

如果 $S_h \leq 0$, 则

$$S_h = \frac{1}{S(A)} \left(\sum A - \sum B \right) \quad (6)$$

使得模板在水平方向滑动,记录使得评价函数值最大的位置即水平方向精确调整后头部的最佳位置。滑动的方向选择,具体可以通过左右两个 B 区域值的相对变化来判断。

“粗一精”双步头部定位算法流程是:首先使用“凸”型模板完成人体头部以及躯干位置的粗定位,然后在粗定位的基础上,分别使用垂直方向精确定位模板和水平方向精确定位模板在垂直向和水平向小范围地滑动搜索,并分别根据两种模板评价函数的函数值精确地定位人体头部。

3.4 算法实现方法

在本算法的实现过程中,涉及了大量在某一个矩形区域内累加计算前景点个数的运算过程。这里采用积分图(integral image)法^[7]加速运算过程,以使该算法能够满足实时要求。

在积分图中,每一点 (x, y) 的值是所有在此点左上部像素的像素值之和,这里使用的是二值化前景区域图像(背景像素为 0,前景像素为 1),所以即每一点的值是所有在此点左上部前景像素的个数之和。积分图 $I(x, y)$ 的定义如下:

$$I(x, y) = \int_{i=0}^x \int_{j=0}^y I(i, j) dj di \quad (7)$$

式中, $I(x, y)$ 为原始图像。在实际数字图像的计算中,采用离散计算式 $\sum \sum I(x, y)$ 完成近似计算。在计算全部 $I(x, y)$ 值的时候,可以把每一点的值都

按照上式计算,但是这样的计算效率很低。为了提高计算 $I(x,y)$ 的效率,采用递推的计算方法^[6]:

$$\forall x,y, I(x,y) = I(x,y) + I(x,y-1) + I(x-1,y) - I(x-1,y-1) \quad (8)$$

递推公式的得出是利用矩形的特性。其可以充分的利用前面已知点的值,提高计算效率。

计算积分图的目的是为了能够更有效地计算出模板内出现前景像素点的个数(即 $\sum A$ 和 $\sum B$ 的值),下面具体阐述其实现过程:

首先通过积分图计算一个矩形区域内前景像素的个数。

如图 6 所示,如果需要计算区域 1 内前景像素的个数,假设此时积分图已经计算得出了。那么在区域 1 的 4 个顶点 A、B、C、D 的值就分别是:

- $I(x_A, y_A)$ = 区域 4 的前景像素个数
- $I(x_B, y_B)$ = 区域 3、4 的前景像素个数
- $I(x_C, y_C)$ = 区域 1、2、3、4 总的前景像素个数

$I(x_D, y_D)$ = 区域 2、4 的前景像素个数

那么由此便可以很容易的得出,区域 1 的前景像素个数可以表示为

$$\sum 1 = I(x_C, y_C) + I(x_A, y_A) - I(x_B, y_B) - I(x_D, y_D) \quad (9)$$

同理,在计算模板内前景像素个数时,只需将模板分成几个矩形块,那么整个模板内前景像素的个数可以通过上述的方法:先分别计算各个矩形块的值然后再求和得到。即例如“凸”型模板内前景像素值的计算公式是(参见图 7):

$$\sum A = I(x_C, y_C) + I(x_A, y_A) - I(x_B, y_B) - I(x_H, y_H) + I(x_E, y_E) + I(x_C, y_C) - I(x_D, y_D) - I(x_F, y_F) \quad (10)$$

通过这样的实现算法,对比直接统计模板内前景

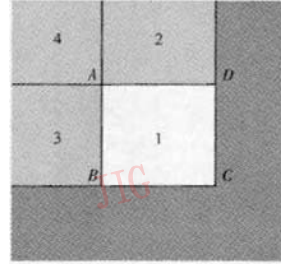


图 6 矩形区域计算图

Fig. 6 The computation integral image of rectangle

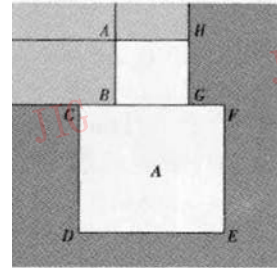


图 7 模板区域计算图

Fig. 7 The computation integral image of mask

点的个数可以大大地简化运算量,达到实时性的要求。

4 头部提取算法实验结果

本算法实现的软件平台是 VC ++ 6.0, 硬件平台是 C4-1.7G, 512RAM 的 PC 机, 实验图像分辨率为 352 × 288。在进行了大量的头部提取实验之后, 实验结果证明了本算法的有效性。图 8 和图 9 是文献[6]与本文提出的“coarse-accurate”双步提取头部算法在人行过程中头部明显偏离躯干中线情况下的算法结果对比。

由图 8(b) 可以看到, 由于人体正在行走中头部略微向前倾, 虽然定位头部的大致的位置是正确的,



(a) 原始输入图像

(b) 文献[6]定位结果

(c) 本文算法定位结果

图 8 转身时人体头部提取结果对比

Fig. 8 Comparison of turning human head detecting results

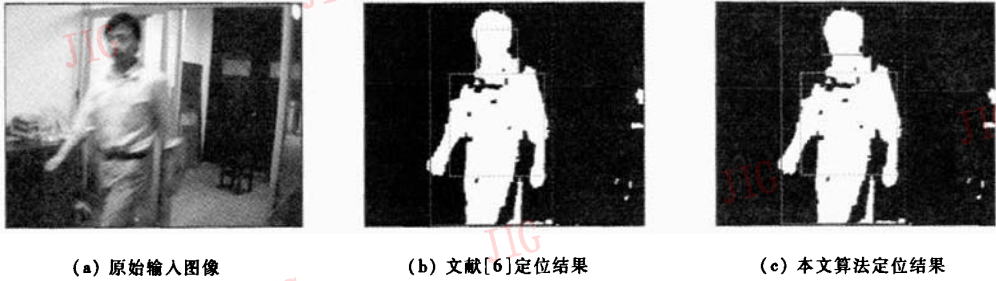


图 9 行走时人体头部定位结果对比

Fig. 9 Comparison of working human head detecting results

但是却不够准确,定位的位置明显靠后。由图 8(c) 可以看到,由于进行了水平向调整,可以非常精确地定位到头部。由图 9(b) 可以看到,由于人体靠镜头较近导致身体比例放大,出现了一段比较长的颈部区域,这使得定位的头部区域实际上在人体的颈部。由图 9(c) 可以看到,由于进行了垂直向和水平向头部位置的调整,故仍然可以很准确地定位到头部区域。

5 结 论

根据人体视觉特征,结合文献[6]算法的优点,本文提出了一种新型的“粗一精”双步头部定位算法,消除了原文献[6]算法对头部和躯干相对位置的严格限制。由实验结果可以看出,“粗一精”双步头部定位算法与文献[6]定位算法相比,在定位精确度上有了很大的提高,该算法能更好地适应复杂情况下人体头部的定位。

下一步的工作可以考虑将水平向和垂直向两个精确定位模板合并成一个独立的精确定位模板;同时在完成第 2 步精确定位时,还可以考虑改变精确定位模板的大小,使算法的鲁棒性进一步增强以适用于更复杂的情况。

参考文献 (References)

- 1 Wren C, Azarbayejani A, Darrel T, et al. Pfindex: real-time tracking of the human body [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, 19(7): 780 ~ 785.
- 2 McKenna S J, Raja Y, Gong S. Object tracking using adaptive colour mixture models [J]. *Lecture Notes in Computer Science*, 1998, 1(1351): 615 ~ 622.
- 3 Ahlberg J. Using the active appearance algorithm for face and facial feature tracking [A]. In: *Proceeding 2nd International Workshop Recognition, Analysis and Tracking of Face and Gestures in Real-time Systems* [C], Los Alamitos, CA, USA: IEEE Comput. Soc, 2001: 86 ~ 72.
- 4 Haritaoglu I, Harwood D, Davis L S. W⁴: A real time system for detecting and tracking people [A]. In: *Proceedings of 3th IEEE International Conference on Automatic Face and Gesture Recognition* [C], Los Alamitos, CA, USA: IEEE Comput. Soc, 1998: 222 ~ 227.
- 5 Zhao Liang. *Dressed Human Modeling, Detection, and Parts Localization* [D]. Pittsburgh, PA: Carnegie Mellon University, 2001.
- 6 Micilotta A, Bowden R. View-based location and tracking of body parts for visual interaction [A]. In: *Proceedings of 13th British Machine Vision Conference* [C], Oxford, United Kingdom: Elsevier Ltd, 2004: 849 ~ 858.
- 7 Viola P, Jones M. Robust real-time object detection [J]. *International Journal of Computer Vision*, 2004, 57(2): 137 ~ 154.