

用于视频头肩序列图像编码的人脸定位方法¹

汪孔桥 沈兰荪 邢 昕

(北京工业大学信号与信息处理研究室 北京 100022)

摘 要 该文运用人脸结构特点和数学形态学提出了一种稳健有效的人脸结构定位方法, 该方法主要用于视频头肩序列图像的编码, 对减小人脸预测误差、提高编码效率起着极为关键的作用。

关键词 MPEG-4 标准, 人脸结构, 数学形态学, 人脸结构定位, 视频编码

中图分类号 TN919.8

1 引 言

随着图像压缩编码技术的飞速发展, 基于图像内容的压缩编码技术受到了广泛关注^[1,2], 并已被 MPEG-4 吸收为关键技术^[3]。在可视电话、视频电视会议、人机交互等领域, 头肩部像普遍存在, 对此, 我们已提出了面向人脸对象的图像编码方法, 即结合人脸检测的小波图像编码方法^[4-6], 获得了较好的压缩效果。对于视频头肩序列图像, 为了实现面向人脸对象的极低比特率视频编码, 应定位出每帧头肩部的人脸结构位置, 使各帧人脸在空间位置上保持一致, 这样将会大大减小人脸的预测误差, 极大地提高视频编码增益, 因此人脸结构定位对头肩序列图像编码效率的提高极为关键, 而现有的一些人脸定位方法只能给出图像中人脸的大致位置^[7-10], 为此, 本文在人脸检测的基础上^[6], 利用人脸的结构特点和数学形态学实现了头肩部的人脸结构定位, 算法具有较强的稳健性和有效性。

2 数学形态学理论

数学形态学是基于形状集合理论而提出的, 其运算本质上是并行的, 便于用并行处理来实现^[11]。用数学形态学算子对图像进行处理, 不需计算每点的局部特征, 方法简单, 故速度快且抗噪声能力比较强。在数学形态学中有四个常用的基本算子: 腐蚀、膨胀、开和闭。运算过程中, 仅有被分析的图像 X 和结构元素 B 参与, 其中 B 的形状是影响运算结果的关键因素。

2.1 腐蚀和膨胀 设 X 和 B 为二维欧氏空间中的两个集合, 对于二值图像, 设 $(X)_b$ 表示 X 平移 b 而得到的图像, 即

$$(X)_b = \{y | y = x + b, x \in X\} \quad (1)$$

则腐蚀和膨胀的定义分别为

$$\text{腐蚀: } X \ominus B = \bigcap_{b \in B} (X)_{-b} = \{y | b \in B, (y + b) \in X\} \quad (2)$$

$$\text{膨胀: } X \oplus B = \bigcup_{b \in B} (X)_b = \{y | y = x + b, x \in X, b \in B\} \quad (3)$$

对于灰度图像, 情况大致类似, 腐蚀和膨胀的定义分别为

$$\text{腐蚀: } X \ominus B = \min_{b \in \Omega} [X(x + b) - B(b)] \quad (4)$$

$$\text{膨胀: } X \oplus B = \max_{b \in \Omega} [X(x - b) + B(b)] \quad (5)$$

其中 Ω 为结构元素 B 的支撑集。

¹ 1998-12-11 收到, 1999-09-31 定稿
国家 863 资助课题

由 (2), (3), (4), (5) 式可以看出, 二值图像的腐蚀和膨胀实质上分别对应求“与”运算和“或”运算, 而灰度图像的腐蚀和膨胀实质上分别对应运算集内求最小值和最大值。

2.2 开和闭 开和闭运算是基于腐蚀和膨胀运算的基础之上的, 开定义为先腐蚀后膨胀, 而闭为先膨胀后腐蚀, 即

$$\text{开: } X \circ B = (X \ominus B) \oplus B \quad (6)$$

$$\text{闭: } X \bullet B = (X \oplus B) \ominus B \quad (7)$$

二值图像和灰度图像的开和闭运算过程相同。一般来说, 开运算将会滤除图像中的孤立噪声点和磨平目标表面的尖峰, 而闭运算则会填平目标表面的凹陷。

3 人脸的细致定位

输入图像经过人脸检测后, 获得了人脸在图像中的大致位置, 但在面向人脸对象的极低比特率视频头肩图像编码中, 需知道人脸的结构位置。以下我们根据人脸的结构特点和数学形态学描述人脸结构定位的方法。

3.1 人脸的水平投影 人脸的眉、眼、鼻孔、嘴区, 从灰度分布上来看, 比周围区域暗, 因此如对面脸作水平投影, 则在眉、眼、鼻孔、嘴的水平位置会分别出现一个极小值, 如图 1(a), 根据各谷点之间存在的位置关系, 可确定眼和嘴等器官的水平位置, 例如:

(1) 如果 x_1, x_2, x_3, x_4 分别对应眉、眼、鼻孔、嘴巴的水平位置, 则 $(3/5)(x_3 - x_1) \leq x_3 - x_2 < 3(x_2 - x_1)$ 且 $(4/7)(x_4 - x_2) \leq x_3 - x_2 < 3(x_4 - x_3)$;

(2) 如果第一个谷点与第二个谷点之间的距离大于第二个谷点与第三个谷点之间的距离, 则第一个谷点被认为是眼睛的水平位置;

(3) 如果第一个谷点与第二个谷点之间的距离小于第二个谷点与第三个谷点之间距离的 $1/3$ 倍, 则第三个谷点被认为是嘴的水平位置。

在实际判决之前, 要去掉一些因噪声干扰而引起的深度较浅的伪谷点, 图 1(b) 给出了一个实际判决的结果。

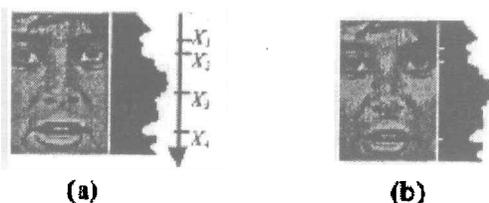


图 1 人脸的水平投影

一旦眼、嘴水平位置确定后, 便可获得这两者之间的垂直距离 h , 令 $w = 0.8 \times h$, 便可确定脸区 (h 和 w 分别对应脸区的高和宽), 其中脸区的上、下边分别落在眼、嘴之上 (如图 2)。

3.2 嘴中心点的确定 欲确定嘴中心点, 必须首先要确定嘴所在的区域, 即嘴区。嘴区可由脸区来确定。本文中把脸区下边中心点作为嘴区中心点, 嘴区的高由鼻与嘴之间的垂直距离来确定, 宽则由脸区的宽来确定, 如图 3。用数学形态学运算对嘴区进行分割处理, 便可确定嘴中心点, 以下为具体的分割步骤 (图 4)。设嘴区原始图像为 I :

(1) 对 I 进行腐蚀运算, 得 $I_e = I \ominus B$;

(2) 用 I 减去 I_e , 取反, 得图像 $I'_e = 255 - (I - I_e)$, 并对之进行二值化得图像 I_b ;

(3) 对 I_b 进行结构闭运算, 得 $I_c = I'_e \bullet B$;

(4) 对 I_c 进行区域标识, 保留最大区, 并求出该最大区的最左端点 (x_1, y_1) 和最右端点 (x_2, y_2) , 则嘴巴的中心点可确定为 (x, y) , 如图 3, 其中 $x = (1/2)(x_1 + x_2)$, $y = (1/2)(y_1 + y_2)$.

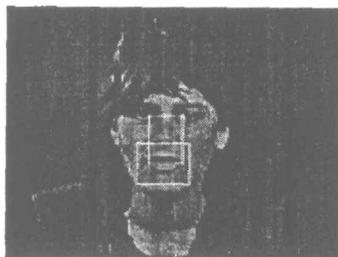


图 2 脸区和嘴区的确定



图 3 嘴中心点的确定



图 4 形态学分割嘴区

3.3 人脸中轴线的求取 尽管人脸形状各不相同, 如方脸、长脸、圆脸等, 但他们一般都具有良好的对称性 (除个别伤残人脸外), 其对称轴为通过双眼中心点和嘴中心点的连线。

设 $F(x, y, \theta)$ 表示一个与脸区同样尺寸的矩形区, 它的下边中心点为 (x, y) , 其中心轴 $f(\theta)$ 与垂直方向的夹角为 θ , 且 $f(\theta)$ 将 $F(x, y, \theta)$ 平分分为 A 、 B 两个区, 如图 5. 如果 (x, y) 点为嘴的中心点, 则一旦 θ 角确定, 人脸中轴线便可确定下来。设 $P_i \in A, P'_i \in B$, 且 P_i, P'_i 关于 $f(\theta)$ 对称, 则 A 、 B 两个区所有对应点的误差绝对值之和为

$$ASE(x, y, \theta) = \int_{P_i \in A, P'_i \in B} |P_i - P'_i| \tag{8}$$

如果人脸的中轴线倾斜角为 θ_f , 则 $ASE(x, y, \theta_f)$ 将达到极小, 即

$$\text{当 } \frac{\partial ASE(x, y, \theta)}{\partial \theta} = 0 \text{ 时, } \theta = \theta_f$$

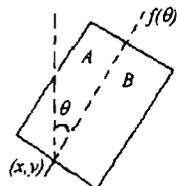


图 5 人脸中轴线的求取



图 6 人脸结构定位

由于用数学形态学运算求得的初始嘴中心点可能与真实嘴中心点存在一个或几个像素距离的误差, 因此需对初始嘴中心点进行调整方能获得较为准确的人脸中轴线, 调整的具体方法为: 将初始嘴中心点向左、右各平移若干像素, 在每平移一次时, 求出一个极小的

$ASE(x, y, \theta)$, 最后在这些极小的 $\{ASE(x, y, \theta)\}$ 中求出最小值, 此时 θ 即为最终的人脸中轴线的倾斜角, 且中轴线经过调整后的嘴中心点, 如图 6。

3.4 人脸结构定位 当人脸中轴线、嘴中心点及眼与嘴之间的垂直距离被确定后, 便可一椭圆标示出该人脸结构的位置, 该椭圆的长轴与人脸中轴线重合 (图 6)。设嘴中心点的坐标为 (x, y) , 中轴线的倾斜角为 θ , 眼到嘴的垂直距离为 h , 则两眼之间的中心点坐标 (x', y') 为

$$\left. \begin{aligned} x' &= x - htg\theta \\ y' &= y - h \end{aligned} \right\} \quad (9)$$

在本文中, 我们选择点 (x', y') 为人脸椭圆的中心, 椭圆的半长轴、半短轴分别为 $a = 1.5 \times h / \cos \theta$, $b = h / \cos \theta$ 。

4 实验结果与结论

本文根据人脸的结构特点和数学形态学运算对人脸结构进行了定位, 算法具有较强的稳健性和有效性。我们对 167 幅人脸姿势各异的图像进行了实验测试, 结果显示: 只要图像的光照均匀, 人脸在图像平面内旋转范围不超过 20° , 向图像深度方向旋转范围不超过 10° , 均能较准确地定位人脸, 图 7 给出了部分实验结果。

该方法主要用于视频头肩序列的编码, 编码时, 由于每帧图像中的人脸位置得到了较准确的定位, 因此各帧图像上的人脸在空间位置上就能得到较好地配准, 编码预测误差将会大大降低, 从而编码效率会得到极大提高。

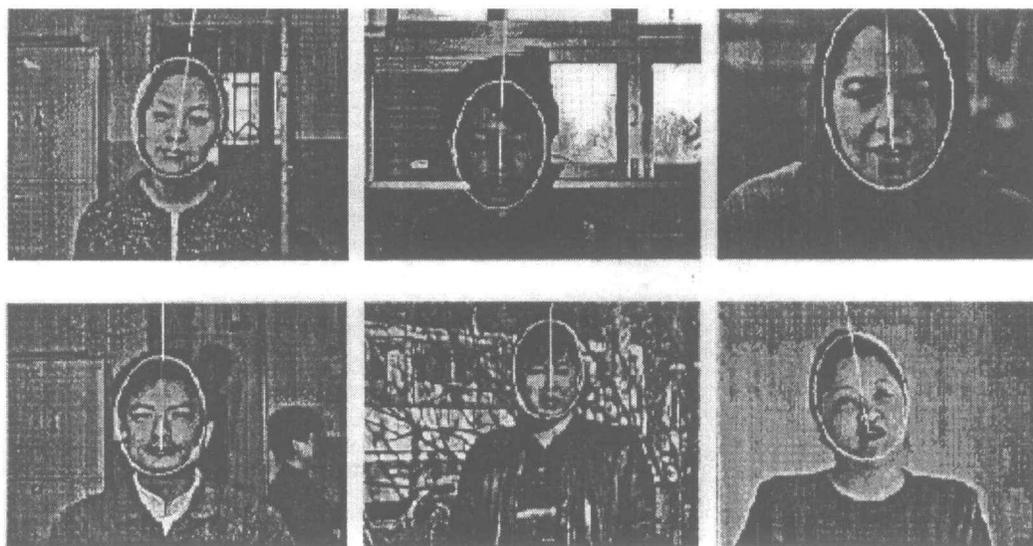


图 7 人脸结构定位的部分实验结果

参 考 文 献

- [1] Gerken P. Object-based analysis-synthesis coding of image sequences at very low bit rates. *IEEE Trans. on Circuits Syst. Video Technol.*, 1994, CASVT-4(1): 228-235.
- [2] Talluri R, Oehler K, Bannon T, *et al.* A robust, scalable, object-based video compression technique for very low bit-rate coding. *IEEE Trans. on Circuits Syst. Video Technol.*, 1997, CASVT-7(1): 221-233.
- [3] MPEG-4 project description, ISO/IEC JTC1/SC29/WG11 N1177, Munich MPEG meeting, January 1996.
- [4] 汪孔桥, 沈兰荪. 一种结合人脸检测的静止小波图像编码方法. *电子学报*, 1998, 6(4): 88-90.
- [5] 汪孔桥, 沈兰荪. 一种结合人脸检测的多码率小波图像编码方法. *高技术通讯*, 1998, 8(9): 6-9.
- [6] Shen Lansun, Wang Kongqiao, Qin Xia, Xing Xin. A multirate wavelet image coding method based on human face detection. *Proc. ICMI'99, Hong Kong: 1999, VI.9-VI.13.*
- [7] Govindaraju V. Locating human faces in photographs. *International Journal of Computer Vision*. 1996, 19(2): 129-146.
- [8] Yang G, Huang T. Human face detection in a complex background. *Pattern Recognition*, 1994, 27(1): 53-63.
- [9] Agui T, Kokubo Y, Nagahashi H, *et al.* Extraction of face regions from monochromatic photographs using neural networks. *Proc. Int. Conf. on Robotics, Singapore: 1992, CV18.8.1-CV18.8.5.*
- [10] Lam K M. A fast approach for detecting human faces in a complex background. *Proc. IEEE ISCAS'98, 1998, Vol.4, 85-88.*
- [11] Pei S-C, Chan F-C. Image sampling structure conversion by morphological filters. *Signal Processing: Image Comm.*, 1994, 6(1): 13-24.

HUMAN FACE LOCATION USED IN
HEAD-SHOULDER VIDEO CODING

Wang Kongqiao Shen Lansun Xing Xin

(Signal and Infor. Processing Lab., Beijing Polytechnic University, Beijing 100022)

Abstract A robust and efficient method of human face fine location is presented using human facial structure and mathematical morphology. The algorithm, which is mainly used in head-and-shoulder video coding, plays a critical role to decrease human face predictive discrepancy and increase coding efficiency.

Key words MPEG-4 standard, Human facial structure, Mathematical morphology, Human face fine location, Video coding

汪孔桥: 男, 1967年生, 博士生, 研究方向为智能化信息处理.

沈兰荪: 男, 1938年生, 教授, 博士生导师, 研究方向为智能化信息处理.

邢 昕: 男, 1974年生, 硕士生, 研究方向为智能化信息处理.