

# 基于小波和 ICA 的面部特征定位<sup>\*</sup>)

王绍宇

(东华大学计算机科学与技术学院 上海 201620)

**摘要** 面部特征的定位是自动人脸识别(AFR)系统的重要组成部分,现有主要包括基于先验知识、几何形状、色彩、外观和关联信息五类方法。本文从信号学的角度,提出了一种基于小波和独立分量分析(ICA)的新方法。先对面部图像进行小波分解,提取出主要代表眼睛和嘴巴特征的水平边缘图像,再把说话人在视频中眼睛的闭合和嘴巴的运动看成是相互独立的运动分量,利用 ICA 分离出眼基和嘴基,然后分别利用它们来重建人脸图像,从而实现眼睛和嘴巴的定位。

**关键词** 面部特征定位,小波分解,独立分量分析

## Facial Feature Localization Based on Wavelet and ICA

WANG Shao-Yu

(School of Computer Science and Technology, Donghua University, Shanghai 201620)

**Abstract** Facial feature localization, an important technique in automatic face recognition, could be classified into five methods, namely ones based on knowledge, geometry information, color, appearance and relative location. Our new method based on wavelet and independent component analysis regards the movements of eyes and mouth as independent components from the view of the signal processing. The wavelet decomposition is first applied to obtain the horizontal edge images. The apparatus basis images, extracted by ICA on these edge images, are then used separately to reconstruct the facial image to localize eyes and mouth.

**Keywords** Facial feature localization, Wavelet decomposition, Independent component analysis

## 1 引言

在基于人脸 3D 模型的极低比特率视频编码和人脸表情识别中,为了获取眼睛和嘴巴等面部器官的特征参数和表情参数,需要对它们进行定位。通常都是利用人脸器官构造的先验知识、面部图像灰度分布特性来对眼睛、鼻子、嘴等器官进行定位。Craw 等利用头形线条模板从低分辨率到高分辨率定位人的头部轮廓,再用线条跟踪器定位眼睛和嘴巴<sup>[1]</sup>,这种方法比较繁琐而且精度也不高。Wong K H 等从灰度图像上提取一维水平或垂直的‘积分投影’曲线,根据曲线的局部极值实现眼睛和嘴巴的定位<sup>[2]</sup>。Brunelli 等利用边缘微分算子求出人脸图像的边缘,再用垂直方向的积分投影曲线用于定位眼睛和嘴巴<sup>[3]</sup>。但是这两种方法容易受光照、人脸姿态以及表情等因素的影响。还有的用神经网络的方法,如多层感知器<sup>[4]</sup>和自组织特征映射<sup>[5]</sup>,先用眼睛和嘴巴的样本图像进行训练,再对人脸图像的每块子区域进行检测定位。这种方法的缺点是需要大量的各种不同尺度、旋转角度和光照条件下的训练样本。

针对有说话人的视频流,为了定位人脸眼睛和嘴巴两个主要表情点,我们提出了一种与现存的根据人脸的五官分布特性以及灰度分布特点来定位人的眼睛和嘴巴截然不同的人脸特征定位方法。

## 2 基于 ICA 的人脸特征定位

我们注意到:为了保持眼睛的湿润,人必须要周期性地眨眼,频率约为 16 次/分钟,而且双眼都几乎在同一时间进行这个眨眼的运动。而且在面向视频的应用中,由于要进行交谈,嘴巴也要进行一系列的开启和闭合的运动,因此我们可以把

眼睛和嘴巴的运动看成相互独立的信号,对这些运动所反映出来的人脸图像进行研究,利用 ICA 来对它们进行盲源分离,从而实现人脸特征的定位。

### 2.1 人脸图像的小波预处理

由于目的是利用视频流中人脸眼睛和嘴巴的运动进行人脸特征的定位,这些运动主要导致眼睛和嘴巴边缘发生变化,因此可以利用小波分解的特点来保留边缘信息,去除人脸的近似信息,从而提高特征定位的准确性。通过 Mallat 等<sup>[6,7]</sup>提出的多分辨率分析可以构造出一对正交镜像滤波器  $h$  和  $g$ ,  $h$  为低通滤波器,  $g$  为高通滤波器。把这两个滤波器分别应用于两个不同的方向可得到应用于 2D 的离散小波变换,从而得到图像的近似图像和细节图像。经过小波分解后,人脸图像的能量大都聚集到了低频部分,也就是近似图像,而人脸的眼睛和嘴巴边缘以及人脸轮廓信息则集中在了高频部分,即细节图像。

用 db1 小波对 akiyo 视频流中的一帧人脸图像进行一次小波分解,得到分解的图像如图 1 所示。

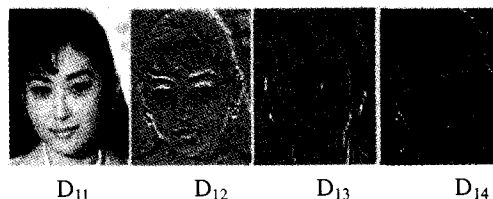


图 1 人脸图像小波一次分解结果

可以看到:在人脸的水平边缘图像  $D_{12}$  (左边第二个图像)中,主要包含了人脸的眼睛和嘴巴的边缘信息,因此我们以  $D_{12}$  为研究对象。

<sup>\*</sup>)国家自然科学基金项目(60271033)。王绍宇 讲师,博士,研究方向:人脸识别与图像分割。

### 2.2 白化处理

白化目的是去除输入图像之间的相关性,得到一白化后的变量  $\tilde{X}$ ,以简化独立分量的提取过程。白化后的  $\tilde{X}$  满足:  $E\{\tilde{X}\tilde{X}^T\}=I$ 。

对于一大大小为  $M \times M$  的人脸边缘图像,把它按逐行展成  $1 \times M^2$  大小的行向量  $X_i$ ,得到训练图像集合为  $X=[X_1^T, X_2^T, \dots, X_n^T]^T$ ,  $n$  为训练样本数目。训练样本的均值定义为:

$$u = E[X] = \frac{1}{n} \sum_{k=1}^n X_k^T \quad (1)$$

$X$  的协方差矩阵定义为:

$$\begin{aligned} \Sigma_x &= E[(X^T - u)(X^T - u)^T] = \frac{1}{n} \sum_{i=1}^n (X_i^T - u)(X_i^T - u)^T \\ &= \frac{1}{n} X X^T \end{aligned} \quad (2)$$

$X$  为  $[X_1^T - u, X_2^T - u, \dots, X_n^T - u]$ , 由于对称矩阵  $\Sigma_x$  的  $\frac{\partial y}{\partial x}$  大小为  $M^2 \times M^2$ , 为了减小计算它的特征值和特征向量的计算量,使用 SVD 作以下处理:

令  $R = X^T X$ ,  $\Lambda$  为由  $R$  的特征值构成的对角矩阵(特征值按从大到小排列),与  $\Lambda$  对应的特征向量矩阵为  $V$ ,可以得到  $\Sigma_x$  的特征向量矩阵的转置为  $\tilde{X} = \Lambda^{-1/2} V^T X^T$ ,  $\tilde{X}$  中的每一行对应于白化处理后的人脸边缘图像。

### 2.3 基于 ICA 的人脸特征定位

人脸边缘图像可认为是对应于眼睛、嘴巴等相互独立的器官基图像(Apparatus Basis Image)线性混和而成的,通过 ICA 方法求出对应眼睛和嘴巴等器官的基图像,人脸边缘图像就从原来的以像素为单位的图像空间表示形式转换到了以相互独立的器官基图像为单位的 ICA 基空间表达形式。要实现眼睛和嘴巴的定位,只需将待定位图像向此空间进行投影,获取相应器官基对应的投影系数,然后利用对应的器官基进行重建即可。

据 ICA 理论:

$$S = W\tilde{X} \quad (3)$$

在这里,  $\tilde{X}$  每行代表了白化处理后的人脸边缘图像,  $S$  的每行代表相互独立的人脸器官基图像,  $W$  为  $k \times n$  大小的分离矩阵,  $k$  为相互独立 ICA 基的数目。

FastICA 是一种基于负熵的独立性判别准则的分离算法,它可以逐个地对独立分量进行分离,并且使用固定点法(fixed point method),通过随机梯度法调节分离矩阵  $M$  来达到优化的目的,从而加快算法收敛的速度。因此我们使用 FastICA<sup>[8]</sup> 算法来对白化后的人脸边缘图像进行器官基的提取。

人脸器官基图像比人脸边缘图像的非高斯性强,因此可以通过分离结果的非高斯性来进行衡量是否完成了对独立的人脸器官基的分离。基于对负熵的近似,非高斯量度的近似公式可以表达为:

$$J(y) \approx c [E\{G(y)\} - E\{G(v)\}]^2 \quad (4)$$

$c$  为常数;  $v$  为均值为零,方差为 1 的高斯变量;  $G$  为非线性函数,我们取为:  $G(y) = y^3$ 。

在独立分量分离过程中,用式(4)对分离结果进行衡量,并对分离矩阵  $W$  进行调整,  $W$  的调整公式为:

$$W(k) = E\{\tilde{X}G'(W(k-1)^T\tilde{X})\} - E\{G'(W(k-1)^T\tilde{X})\}W(k-1) \quad (5)$$

每次迭代后都要对  $W(k)$  作归一化处理:  $W(k) = W(k) /$

$|W(k)|$ ,以保证分离出来的独立分量有单位能量。在提取出某个独立分量后,都要从输入信号  $\tilde{X}$  中减去这个独立分量,再进行下一个独立分量的提取,如此重复进行直到提取出所有的独立分量。

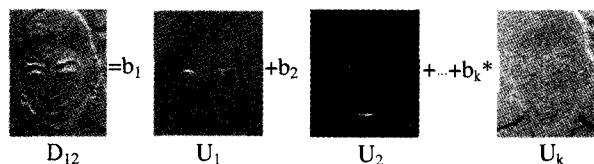


图2 脸边缘图像器官基表示( $u_1, u_2, \dots, u_k$  代表相互独立的器官基,  $b_1, b_2, \dots, b_k$  代表对应的器官基的系数)

试验结果如图 2 所示,可以看到通过 ICA 可以得到了一系列与眼睛、嘴巴和人脸轮廓等对应的 ICA 器官基。要实现对眼睛和嘴巴的定位,只需在重建的过程中保留对应的 ICA 器官基,把其他的 ICA 器官基都置为零,就可以通过重建图像实现目标特征的定位。

### 3 试验结果

实验采用标准视频流 akiyo(30f/s)来对本文提出的算法进行测试。定位过程主要分为 3 步:

第一:首先获取背景图像,在人脸进入视频窗口开始说话后,利用帧差和混和肤色模型对人脸进行检测,获取包含人脸的大致的矩形区域。

第二:在视频开始的前 6 秒中自动随机抽取其中的 50 帧作为训练样本,利用小波分解过滤掉人脸的近似信息,得到人脸的水平边缘图像,再使用 FastICA 算法提取出独立分量,获取与嘴巴和眼睛等器官对应的 ICA 器官基。

第三:把待定位的图像先进行人脸检测,再利用小波分解得到对应的边缘图像,将此图像投影到由所有独立分量张成的特征空间中得到相应的投影系数,最后将与待定位器官无关的 ICA 器官基都置为零,进行人脸图像的重建,从而得到突出表征所求器官位置的重建图像。

最后对得到的重建图像进行简单的阈值处理就可以得到特征目标的位置信息。图 3 的试验结果表明我们提出的这种新的方法完全可以对视频流中眼睛和嘴巴进行有效的定位。

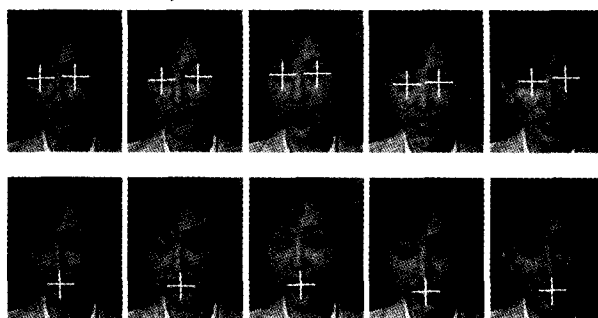


图3 akiyo 测试视频流中眼睛和嘴巴的定位结果

**结论** 本文在人脸特征的定位上,提出了不同于传统的根据人脸器官灰度经验分布的特性来定位人脸特征的一种新方法。把视频流中眼睛和嘴巴运动所分别造成的人脸图像的变化看成是相互独立的信号,利用人脸检测和小波分解去掉不感兴趣的背景和人脸的近似部分,获取代表眼睛和嘴巴等

(下转第 261 页)

部分却不尽相同。DCPI 捕获关于硬件事件的信息,例如 cache 不命中和分枝预测错误等,而 TIPME 捕获的信息是 GUI 事件和操作系统状态的转换。由于它们关注的目标不一样,所以最终得到的结果也不一样。TIPME 着重于确认和补救操作系统中引起用户感知性能问题出现的部分,而 DCPI 的重点是得知当前硬件在系统执行时对性能的影响。

#### 4.2.3.4 ARM

应用程序反应测量(Application Response Measurement, 简记为 ARM)直接地测量了应用程序的反应时间。它是通过让客户端程序在操作前和后分别调用一次 API 函数得到反应时间的结果。现在,很多大型、内部或商业的应用程序都已利用 ARM 的 API 进行测试,而在不久的将来,更多日常用到的普通应用程序也会利用这个 API 协助进行测试。

#### 4.2.3.5 其它研究

Endo 等人在 MS-Windows 的客户端和服务器间通过 CPU 的活跃和信息交换来推断反应时间。Cota-Robles 和 Held 也认为只靠吞吐量不足以代表计算机系统迅速处理交互式用户请求以及其他实时服务,例如视频和音频等回放的能力。他们关心的是 Windows NT 和 Windows 98 操作系统对于处理实时工作的能力,测量了这两个系统传送硬件中断到相关处理器的速度和可靠性<sup>[7]</sup>。Windows NT 和 Windows 98 有一个共享的视窗驱动模型(Windows Driver Model)。在这个模型内,传统上由中断处理器处理的计算会改由内核线程或通过延迟过程调用(Deferred Procedure Calls)完成。它们都要求操作系统进行调度决定,以及当操作系统遇到较长的中断传送延迟时要作出适当的处理。实验表明,利用吞吐量的基准测试结果不能够正确地表现出实时交互系统的性能。虽然 Windows NT 比起 Windows 98 能够提供至少一个数量级更好的实时反应,但在基于吞吐量的基准测试(Win-stone benchmark)中显示的结果,这两个系统的得分相差却不超过 10 到 20 个百分点。

#### 4.2.4 用户感知性能测试存在的问题

当前有些测试目标尽管是交互性能,但也仅仅限于应用软件的使用性能。测试在静态的系统环境下进行,这不能动态反映系统资源变化对交互性能的影响。另外,利用 CPU 空闲区间或事件集合状态间接获取延迟时间的测试手段容易受进程调度和 I/O 量大等行为复杂事件的影响,不适合测试延迟时间长的交互式事件。这些问题导致测试结果不能有效指导设计人员从最终用户的角度同时优化系统与应用。

目前,针对事件处理延迟,最大的问题在于如何确定  $T$  和  $\beta$  值。由于它们涉及到人的心理学,现在只能通过收集大量的用户统计得出一个平均值。用户感知性能测试最终的度量指标都会包含有用户的主观因素,只有能够找到合适的方法对这些主观点进行定义和赋值,才可以更有效和准确地对桌面操作系统和交互式系统进行性能的度量。

**结束语** 从桌面操作系统的性能测试技术来看,日后的重点还应放在基于吞吐量和基于用户感知性能两种测试方法的研究上。目前比较广泛使用的基准程序都是基于吞吐量测试的方法。其特点是容易进行测试,而且能有效地表现出某些系统的性能,例如科学计算或编译过程等。但缺点是它能提供的信息太少,不能反映桌面操作系统的特性并且有效地测试交互式系统的性能。而基于用户感知性能的测试,可以更真实地反映用户在使用桌面操作系统时的性能。但目前没有很完善的方法进行测试,现在主要是依靠连续监控技术和事后分析进行度量。

桌面操作系统性能测试研究现状中另一个明显不足是理论研究 and 工程应用相脱节。日后应该进一步改善这两种测试方法目前的不足。还有更重要的一点是结合这两种测试方法,使开发出来的性能测试基准程序能够更有效地满足应用领域的需要。

## 参 考 文 献

- 1 Chen JB, Endo Y, et al. The measured performance of personal computer operating systems. ACM Transactions on Computer Systems, 1996, 3~40
- 2 McVoy LW, Staelin C, Imbench; Portable Tools for Performance Analysis. In: Proceedings of the 1996 USENIX Technical Conference, San Diego, CA, 1996
- 3 Brown AB, Seltzer MI. Operating system benchmarking in the wake of Imbench; A Case Study of the Performance of NetBSD on the Intel x86 Architecture. In: Proceedings of the 1997 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems, Seattle, Washington, United States, 1997
- 4 Lee DC, Crowley PJ. Execution Characteristics of Desktop Applications on Windows NT. In: Proceedings of the 25th Annual International Symposium on Computer Architecture, Barcelona, Spain, 1998
- 5 Endo Y, Wang Z, et al. Using latency to evaluate interactive system performance. In: Proceedings of the Second USENIX Symposium on Operating Systems Design and Implementation, Seattle, Washington, United States, 1996
- 6 Endo Y, Seltzer M. Improving Interactive System Performance Using TIPME. In: Proceedings of the 2000 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems, Santa Clara, California, United States, 2000
- 7 Cota-Robles E, Held JP. A comparison of Windows driver model latency performance on Windows NT and Windows 98. In: Proceedings of the Third Symposium on Operating Systems Design and Implementation, New Orleans, Louisiana, United States, 1999

(上接第 200 页)

器官的边缘细节信息,再利用 ICA 的方法提取出相互独立的 ICA 器官基,然后只用对应的器官基来进行图像重建,从而达到对人脸特征的定位。此方法不适用于静态和没有表情的人脸,但是对于针对存在面部表情的低比特率视频编码和表情识别非常有效。下一步的研究工作将是对 ICA 算法进行改进,提高收敛的速度,期望能够达到实时的特征定位。

## 参 考 文 献

- 1 Craw I, Tock D, Bennett. Finding face features. In: Proc. of the Second European Conference on Computer Vision, 1992. 92~96
- 2 Wong K H, Law H M, Tsang P W M. A system for recognizing human faces. In: Proceedings of the International Conference on Acoustics, Speech and Signal Processing, 1989. 1638~1642

- 3 Brunelli R, Poggio T. Face recognition: features versus templates. IEEE Transaction on Pattern Analysis and Machine Intelligence, 1993, 15(10): 1042~1052
- 4 Vincent J M. Face finding in images. Application of Neural Network. In: Murray N F, ed., 1995. 35~70
- 5 Takacs B, Wechsler H. Locating facial features using SOFM. ICPR1994 B, 55~60
- 6 Bartlett M S, Lades H, Sejnowski T J. Independent component representations for face recognition. In: Proc. SPIE Conf. Human Vision. Electronic Imaging III, Vol 2399, 1998. 528~539
- 7 Mallat S. A theory of multi resolution signal decomposition: The wavelet representation. IEEE Transaction on PAMI, 1989, 11: 674~693
- 8 Bingham E, Hyvärine A. A fast fixed-point algorithm for independent component analysis of complex valued signals. International Journal of Neural Systems, 2000, 1: 1~8