

# 基于模板匹配与人工神经网络确认的人脸检测

梁路宏,艾海舟,徐光佑,张 钊

(清华大学计算机系,智能技术与系统国家重点实验室,北京 100084)

**摘 要:** 本文提出了一种基于模板匹配与人工神经网络确认的人脸检测算法. 算法使用一组双眼-人脸模板对搜索候选人脸,并在搜索过程中使用多层感知器进行确认,以减少错误报警. 大量实验证明了该算法的有效性.

**关键词:** 人脸检测; 模板匹配; 多层感知器

**中图分类号:** TP391.4      **文献标识码:** A      **文章编号:** 0372-2112 (2001) 06-0744-04

## Face Detection Based on Template Matching and Neural Network Verification

LIANG Lu-hong, AI Hai-zhou, XU Guang-you, ZHANG Bo

(Dept. of Computer Science and Technology, Tsinghua University, State Key Laboratory of Intelligent Technology and Systems, Beijing 100084, China)

**Abstract:** In this paper a face detection algorithm based on template matching and neural network verification is presented. Two types of templates: eyes-in-whole and face itself, are used one by one in template matching for searching face candidates. A three layer perceptron is used in the template matching procedure to verify each face candidate to exclude most of the false alarms. Experimental results demonstrate the feasibility of this approach.

**Key words:** face detection; template matching; multi-layer perceptron

### 1 引言

人脸检测是指在输入图像中确定所有人脸的位置、尺度和位姿. 人脸检测的研究来源于实际问题, 在出入的安全检查、视觉监测、智能人机接口、基于内容的检索、数字视频处理等诸多领域有着极为广泛的应用价值. 人脸检测已经成为近年来计算机视觉与模式识别领域的研究热点之一.

人脸是一类自然形体, 虽然具有很强的共性, 但由于个体的外貌、表情以及光照等因素的影响, 使得人脸图像具有比较复杂而细致的模式变化. 因此, 人脸检测是一个极具挑战性的模式分类问题. 如果能够比较成功地构造出人脸检测系统, 将为解决其它类似的复杂模式检测问题提供有意的启示, 因而人脸检测的研究具有十分重要的学术价值.

目前已有的脸检测方法主要有两类: 快速检测算法有基于肤色的方法<sup>[1]</sup>、基于马赛克规则的方法<sup>[2]</sup>等; 基于模板的方法有基于特征脸的方法<sup>[3,4]</sup>、基于聚类学习的方法<sup>[5]</sup>、基于人工神经网络的方法<sup>[6]</sup>以及基于支持向量机(SVM)的方法<sup>[7]</sup>等. 总体而言, 快速检测方法适用于场景受到较强约束的图像中的人脸检测; 而基于模板的诸方法能够适用于一般场景图像中的人脸检测, 有较强的鲁棒性, 具有很大的发展潜力.

本文提出了一种基于模板的方法, 用于检测一般场景图像中的正面端正人脸. 与已有的方法不同, 该算法将直接的模板匹配与人工神经网络确认相结合, 并使用模板匹配检测与“自

举”(bootstrap)方法相结合的手段收集人工神经网络的训练样本. 大量实验证明了该方法的有效性: 对我们自己建立的一般场景图像库(123幅图像、398个人脸)的检测正确率达到了87.9%, 对相关文献中常用作对比的MIT图像库(20幅图像、132个人脸)<sup>[5]</sup>的检测正确率也达到了75.9%.

### 2 检测算法

本文算法的框架是: 穷举搜索输入图像中所有可能尺度和位置的矩形区域(以下称为检测窗口), 对每个窗口采用直接的模板匹配与人工神经网络确认相结合的方法进行检测. 检测窗口的处理步骤如图1所示: 首先使用双眼模板匹配进行粗筛选; 然后对窗口内图像进行梯度照度修正<sup>[5,6]</sup>和均方差标准化<sup>[8]</sup>, 消除光照变化的影响; 接下来使用不同长宽比的人脸模板进行匹配, 若匹配度超过阈值, 则将窗口区域重采样为20×20输入人工神经网络进行确认; 若通过确认, 则加入候选人脸列表.

#### 2.1 模板匹配

采用多个人脸样本取平均的方法构造检测用的双眼-人脸模板对. 选取50幅标准证件照, 手工划出人脸区域作为人脸样本, 将其尺度、均值及方差标准化后取灰度平均得到平均人脸图像, 进一步重采样为20×20的原始模板. 拷贝原始模板的双眼部分, 进行灰度分布标准化后作为双眼模板; 对原始

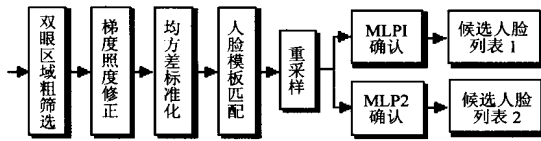


图 1 对检测窗口的处理步骤



图 2 双眼-人脸模板对

模板分别按照 1 0.9、1 1.1、1 1.1、1 1.2 和 1 1.3 的宽长比变形为  $20 \times 18$ 、 $20 \times 20$ 、 $20 \times 22$ 、 $20 \times 24$  和  $20 \times 26$ ，以适应不同形状的人脸，对它们进行灰度分布标准化后作为人脸模板（参见图 2）。

本文使用了一种混合匹配准则。设  $T[M][N]$ 、 $\mu_T$ 、 $\sigma_T$  和  $R[M][N]$ 、 $\mu_R$ 、 $\sigma_R$  分别为人脸模板和检测窗口图像的灰度矩阵、灰度均值与方差，则二者的相关系数  $r(T, R)$  和对应像素灰度值的平均偏差  $d(T, R)$  分别为：

$$r(T, R) = \frac{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (T[i][j] - \mu_T)(R[i][j] - \mu_R)}{M \cdot N \cdot \sigma_T \cdot \sigma_R} \quad (1)$$

$$d(T, R) = \sqrt{\frac{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (T[i][j] - R[i][j])^2}{M \cdot N}} \quad (2)$$

$r(T, R)$  和  $d(T, R)$  分别表征了输入图像区域与模板间图像向量的夹角和欧氏距离，综合考虑这两个方面的因素，对  $d(T, R)$  进行数学处理后与  $r(T, R)$  加权求和作为匹配程度的准则：

$$D(T, R) = r(T, R) + \frac{1}{1 + d(T, R)} \quad (3)$$

其中  $\alpha$  为权重系数，取经验值  $\alpha = 35.0$ 。

## 2.2 MLP 确认

使用三层感知器 (Three Layer Perceptron, 简称 MLP) 对通过模板匹配的检测窗口进行确认。MLP 有 400 个输入层节点，对应于  $20 \times 20$  图像窗口中按行展开的各个像素；隐层节点与输入层和输出层节点全连接；输出层节点一个，输出范围为  $[0, 1]$ ，理想状态下 1 代表“人脸”，0 代表“非人脸”；MLP 采用了 Sigmoid 非线性函数。

基于神经网络的人脸检测方法中“非人脸”样本的选取是一个难题。目前的解决方法是采用“自举”(bootstrap) 算法<sup>[6]</sup>，具有很大的盲目性。因此本文采用了模板匹配检测与“自举”算法相结合的“非人脸”样本选取方法，减少了盲目性。方法如下：

(1) 在若干训练用人脸图像中手工标出 548 个人脸样本，利用双眼和嘴角特征进行头部倾斜角度、脸部长宽比和尺度的标准化。然后经过左右 5 度旋转、左右镜像以及灰度分布标准化，最终得到 3288 个“人脸”样本。

(2) 选取 42 幅不同类型的包含和不包含人脸的图像，使用单纯的模板匹配方法<sup>[8]</sup>进行检测，收集错误报警构成初始的“非人脸”样本集。

(3) 使用步骤 1 得到的“人脸”样本集和已有的“非人脸”样本集训练 MLP。

(4) 使用模板匹配与 MLP 确认相结合的方法检测步骤 2 中的 42 幅图像，收集错误报警，随机选取一部分（约 50%）作为新的“非人脸”样本。

(5) 在已有的“非人脸”样本集中随机选取一部分（约 70%），与步骤 4 新收集的样本共同构成新的“非人脸”样本集。从步骤 3 重复进行训练和收集，直至收集到足够的“非人脸”样本用于训练。

本文使用隐节点数分别为 32 和 36 的两个 MLP（以下简称 MLP1 和 MLP2），分别独立地进行训练、检测和收集样本的迭代过程。两个 MLP 使用了相同的“人脸”样本，但通过使用不同的模板匹配参数生成不同的初始“非人脸”样本，使得两个迭代过程收集的“非人脸”样本具有各自的偏向性。经过上述循环，MLP1 收集了 2802 个“非人脸”样本，MLP2 收集了 3077 个“非人脸”样本，用于最终的训练。

## 2.3 搜索算法

为了解决多尺度人脸的检测问题，将输入图像以 1.2 的比率不断重采样缩小，在缩小后的图像中逐像素移动检测窗口进行匹配和验证。为了准确定位人脸，以模板匹配程度为相似性度量，采用在局部范围内搜索最佳匹配窗口的方法得到人脸的位置。具体而言，就是建立一个候选人脸列表，在搜索过程中将“人脸”窗口逐个加入，如果该窗口与表中已有的候选人脸相重叠，则保留二者中匹配程度高的。如图 1 所示，我们为两个 MLP 分别建立候选人脸列表，在搜索过程中各自加入通过了相应的 MLP 确认的区域，从而最终得到两组模板匹配加单个 MLP 确认的检测结果。

## 2.4 检测结果的仲裁

如图 3 所示，由于两个 MLP 在网络结构和“非人脸”训练样本方面的差异，在它们所对应的两个人脸列表中正确检测到的人脸在位置、尺度上几乎完全一一对应，而错误报警则大多不同。因而可以根据位置和尺度对两个列表中的候选人脸进行仲裁，即只保留在位置和尺度上近似重叠的候选人脸，以减少错误报警。虽然经过仲裁，检测正确率会有所降低，但错误报警率降低得更多，总体而言检测性能将有所提高。



图 3 检测结果的仲裁 (A) MLP1 确认的结果；(B) MLP2 确认的结果；(C) 仲裁的结果

## 3 实验结果

在两个与训练集完全分离的测试集上进行检测实验。测试集 A 是作者建立的图像库，共 123 幅具有复杂背景的图像，包含 398 个正面端正的人脸。测试集 B 取自 MIT 的人脸检测测试集<sup>[5,6]</sup>，共 23 幅图像。去除其中的线画图，选取其余 20 幅包含 132 个正面端正人脸的图像用于测试。

分别使用四种不同的方法对测试集进行检测. 实验中取模板匹配度的阈值为 0.9, 两个 MLP 的阈值均为 0.9. 假设图像中人脸的个数、尺度未知, 采用多尺度搜索的方法, 每种方法各检测了 57, 283, 473 个图像窗口. 表 1 是实验结果统计, 其中错误报警率是错误报警数除以检测窗口数所得的数值.



图 4 测试集 A 的部分实验结果. (A) 扫描的普通照片; (B)、(C) 摄像机拍摄的图片; (D) Internet 下载的图片

可以看到, 模板匹配方法增加 MLP 确认后, 不仅错误报警数大为减少, 而且检测正确率也有较大提高 (第 4 节将进一步说明); 双 MLP 仲裁的方法使得错误报警进一步减少, 而检测正确率只是略有损失. 图 4 是部分实验结果. 本文算法具有较快的检测速度. 以图 4B 为例, 图像大小为  $352 \times 288$ , 检测需搜索 251, 593 个图像窗口, 使用方法 2 和方法 4 的处理时间分别为 57.0 秒和 57.5 秒 (PIIF450 CPU, 128M 内存).

表 2 给出了本文算法对测试集 B 的检测结果统计, 以及文献 [5~7] 中所列的统计数据. 可以看到, 本文的方法取得了可以比较的结果, 但仍略逊于最好的结果. 主要的原因是测试集 B 中有几幅模糊的图像, 其中的人脸区域无法通过检测时的双眼区域粗筛选. 若排除两幅最模糊的图像 Arsenal 和 Usa, 对其余图像的检测正确率可以达到 80.8% (使用 MLP1 进行确认). 图 5 是部分实验结果.

表 2 对测试集 B 的检测实验统计

实验方法	使用的图像	正确检测	正确率	错误报警
方法 2. 模板匹配 + MLP1 确认	20 幅图像 (132 个人脸)	100	75.8%	78
方法 3. 模板匹配 + MLP2 确认		98	74.2%	85
方法 4. 模板匹配 + MLP1、MLP2 确认 + 仲裁		95	72.0%	27
聚类学习方法 (MIT) [5]	23 幅图像 (149 个人脸)	119	79.9%	5
神经网络方法 (CMU) [6]	23 幅图像 (155 个人脸)	131	84.5%	8
SVM 方法 [7]	23 幅图像 (155 个人脸)	115	74.2%	20

#### 4 实验分析

实验表明, MLP 确认可以提高检测正确率. 为了说明这一现象, 对一幅测试图像的局部进行重采样压缩, 使得人脸的尺度与模板的尺度相同, 分别使用模板匹配和 MLP1 对该图像进行“卷积”; 同时分别使用第 3 节中的算法 1、2 进行检测 (如图 6).

表 1 对测试集 A 的检测实验统计 (123 幅图像, 包含 398 个人脸)

实验方法	正确检测	正确率	错误报警	错误报警率
方法 1. 模板匹配	305	76.6%	4837	$8.44 \times 10^{-5}$
方法 2. 模板匹配 + MLP1 确认	350	87.9%	293	$5.11 \times 10^{-6}$
方法 3. 模板匹配 + MLP2 确认	353	88.7%	375	$6.55 \times 10^{-6}$
方法 4. 模板匹配 + MLP1、MLP2 确认 + 仲裁	342	85.9%	132	$2.30 \times 10^{-6}$

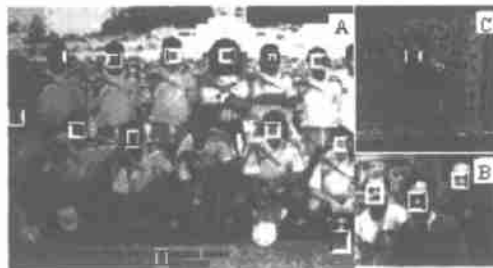


图 5 测试集 B 的部分实验结果

如图 6(B)、(C) 所示, 模板是比较粗略的人脸模型, 某些“非人脸”区域与它的匹配度可能会超过人脸区域与它的匹配度. 如果这些“非人脸”区域与人脸区域相重叠, 就会使真实的人脸被“淹没”掉. 而 MLP 经过了各种类型正反样本的训练, 绝大多数人脸区域都对应着一个孤立的局部极大输出 (图 6(D)). 当然 MLP 也会在某些“非人脸”区域产生极大输出, 但由于它与人脸模板是两种不同的模型, 产生相同错误的可能性很小, 所以经过 MLP 确认, 可以排除掉大部分模板匹配的误差局部最大点, 从而使被“淹没”的一些人脸被检测到 (图 6(E)、(F)).

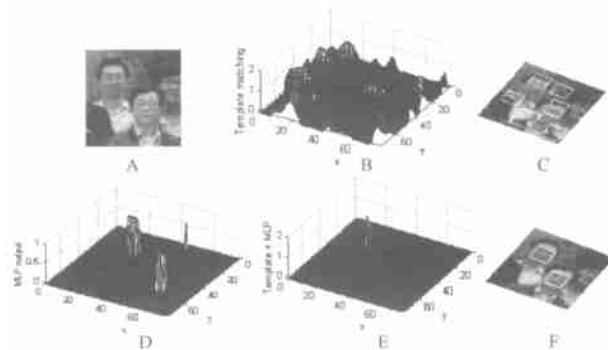


图 6 对图像进行“卷积”和检测的结果. (A) 原图; (B) 模板匹配“卷积”的结果; (C) 模板匹配检测的结果; (D) MLP1“卷积”的结果; (E) 使用 MLP1 确认的模板匹配“卷积”并取阈值的结果; (F) 模板匹配 + MLP1 确认的检测结果

同时, 使用模板匹配环节对于检测和 MLP 训练的益处也是明显的. 多数情况下, 人脸模板匹配的输出能够反映图像区域与通常意义上的人脸的相似程度, 因而可以通过搜索局部最优匹配点来定位人脸. 这种搜索方法相对于单纯的神经网络方法中取“质心”的方法 [6] 具有更加明确的意义. 而且, 根据测试集 A 的实验数据, 只有约 3% 的图像窗口通过了模板匹配环节, 需要进行 MLP 确认, 而模板匹配的计算量远远小于 MLP, 所以本文的方法与单纯的神经网络方法相比具有更快的检测速度. 另一方面, 如果将模板匹配看作是 MLP 检测的粗

筛选环节,它的作用则在于大大缩小了 MLP 所面对的特征空间,从而降低了 MLP 训练的难度,尤其是“非人脸”样本选取的难度。

## 5 总结

本文针对一般场景图像中的正面端正人脸的检测问题,提出了一种基于直接的模板匹配与人工神经网络确认相结合的方法。在 MLP 的训练中,使用了模板匹配检测与“自举”方法相结合的手段收集训练样本,减少了“非人脸”样本选取的盲目性。实验表明,该方法在使用较少训练样本的情况下,构造出了性能较高的人脸检测系统。

本文算法与其它基于样本学习的人脸检测方法一样,检测性能也与训练样本的选取有一定的关系。通过增加训练样本的类型、数量,本文构造的人脸检测系统的性能可以得到进一步的提高。

## 参考文献:

- [ 1 ] C Garcia and G Tziritas. Face detection using quantized skin color regions merging and wavelet packet analysis [J]. IEEE Trans., 1999, Multimedia-1 (3):264 - 277.
- [ 2 ] G Z Yang and T S Huang. Human face detection in a complex background [J]. Pattern Recognition, 1994, 27(1):53 - 63.
- [ 3 ] B Moghaddam and A Pentland. Probabilistic visual learning for object representation [J]. IEEE Trans., 1997, PAMI-19(7):696 - 710.
- [ 4 ] B Moghaddam and A Pentland. Beyond linear eigenspaces: bayesian matching for face recognition [A]. In: Wechsler H et al, eds. Face Recognition from Theory to Applications [C]. New York, 1998:230 - 243.
- [ 5 ] K Sung and T Poggio. Example-based learning for view-based human face detection [J]. IEEE Trans., 1998, PAMI-20(1):39 - 51.
- [ 6 ] H A Rowley, S Baluja, et al. Neural network-based face detection [J]. IEEE Trans., 1998, PAMI-20(1):23 - 38.
- [ 7 ] E Osuna, R Freund, et al. Training support vector machines: an application to face detection [A]. Proc. of CVPR '97 [C], Puerto Rico, 1997:130 - 136.
- [ 8 ] 梁路宏,艾海舟等.基于多模板匹配的单人脸检测[J].中国图像图形学报,1999,4A(10):823 - 830.

## 作者简介:



梁路宏 男. 1973 年 9 月生于河北. 1992 年毕业于清华大学计算机系, 获学士学位. 同年攻读该系计算机应用专业博士学位. 研究方向为计算机视觉、模式识别.



艾海舟 男. 1964 年 1 月生于江苏丹阳. 1985、1988、1991 年分别获得清华大学学士、硕士、博士学位. 1994.9 ~ 1996.8 在比利时布鲁塞尔自由大学作博士后研究, 现为清华大学计算机系副教授. 研究方向为计算机视觉、模式识别.