

基于级联神经网络的人脸检测方法的研究*

陈泽宇 戚飞虎

(上海交通大学计算机科学与工程系, 上海, 200030)

TP391.41

摘要:提出一种基于颜色信息与多级神经网络的人脸检测方法,通过把基于不同颜色分量的多级神经网络级联起来,采用彩色图像的亮度分量和色度分量作为级联神经网络的分类特征,从而有效提高了人脸检测的正确率,同时也加快了人脸检测速度。

关键词:人脸检测,级联神经网络结构,彩色模型,颜色分量。

图像处理

CASCADING NEURAL NETWORKS-BASED HUMAN FACE DETECTION*

CHEN Ze-Yu QI Fei-Hu

(Department of Computer Science and Engineering, Shanghai Jiaotong University, Shanghai 200030, China)

Abstract A new face detection method based on color information and multi-stage neural networks was proposed. The system utilizes brightness and chrominance components of color images as the features of neural network classifiers, and cascades multiple color components-based neural networks. Experimental results show that this method can detect human faces effectively and fast.

Key words human face detection, cascading neural network architecture, color models, color components.

引言

人脸图像的自动识别,长期以来一直是图像处理与模式识别领域的一个重要研究课题,在个人辨识、访问控制、人机交互、远程电视会议、基于内容的图像检索等方面有许多应用。人脸检测是完成自动人脸识别任务的第一步,是人脸识别的基础,目前主要有以下几种方法:相关模板法^[1]、可变形模板法^[2]、特征组法^[3]、神经网络法^[4],还有些方法运用图像的颜色信息或运动信息进行人脸检测^[5,6]。

本文提出一种基于颜色信息与多级神经网络的人脸检测方法,通过把基于不同颜色分量的多级神经网络级联起来,采用彩色图像的亮度分量和色度分量作为级联神经网络的分类特征,有效提高了人脸检测的正确率,加快了人脸检测速度。

1 系统描述

整个人脸检测系统包括尺度变化与窗口扫描部分和级联的神经网络部分。

1.1 尺度变化与窗口扫描部分:

由于图像来源的多样化,图像中人脸的个数、大小往往各不相同,本方法采用基于金字塔的子采样方法,通过对整幅彩色图像进行尺度变化和窗口扫描,切割,解决图像中待测人脸的位置、大小问题。

系统首先对整幅输入图像按比例逐级缩小(尺度变化),而在每一级尺度变化中,按照从上到下、从左到右的窗口扫描方案,以一个 20×20 的窗口模式扫描、切割缩小后的图像,最后把窗口切割得到的彩色窗口图像送入级联的神经网络进行分类判别,这个过程循环往复,直到尺度变化后的图像小于窗口大小为止。

1.2 级联的神经网络部分:

为了判断彩色窗口图像是不是一张人脸,系统中的神经网络部分由多个级联的神经网络结构所组成,每一个神经网络均基于一种颜色分量,被上一级神经网络拒识的彩色窗口图像将被送给下一级神经网络,用另一种颜色分量进行判别,由此,系统可以根据各级神经网络的输出,综合判断出彩色窗口图像是不是一张人脸。

* 国家自然科学基金(编号:69772002)资助项目
稿件收到日期 1999-09-12,修改稿收到日期 1999-10-12

* The project supported by the National Natural Science Foundation of China
Received 1999-09-12, revised 1999-10-12

在每一级神经网络中,彩色窗口图像首先被转换成指定的颜色分量图像,该颜色分量图像接着经过一个人脸掩膜处理过程,滤除图像中人脸轮廓以外的背景成分,然后,系统对掩膜处理后的颜色分量图像进行图像增强处理(如直方图均衡),只可能扩大不同类别图像模式(即人脸与非人脸)间的距离,最后,该窗口图像被送入训练好的神经网络进行人脸判别,若是人脸则系统将把该人脸在原始图像中的位置和大小信息输入,若该窗口图像被本级神经网络所拒识,系统将把原彩色窗口图像提供给基于其它颜色分量的下一级神经网络作进一步的判别。

2 神经网络及其级联

2.1 彩色模型与颜色分量

在对彩色视觉的研究过程中,为了用定量方法描述颜色信息,人们先后提出了许多种彩色模型,我们采用了其中的3种:(1) RGB彩色模型:用于显示和保存彩色图像的最常用的彩色模型,由R(红)、G(绿)、B(蓝)三个色度分量组成;(2) HSV彩色模型:建立在人眼视觉基础上的彩色模型,包括H(色调)、S(饱和度)、V(亮度)3个分量;(3) YES彩色模型: Xerox公司提出的一种彩色模型,包括亮度分量Y和色度分量E和S。在这些彩色模型中,HSV和YES模型都把亮度信息与色度信息分离开来,便于图像增强等彩色图像处理工作,而RGB模型的3个色度分量则在包含颜色信息的同时也包含了大量的亮度信息。

通过一定的转换公式^[7],可以把彩色图像从RGB模型转换成其它彩色模型,从而获得对应的颜色分量图像,这些颜色分量由于来源、定义、算法各不相同,因此具有不同的特性,可以在基于级联神经网络的分类判别中发挥作用。

式(1)是RGB模型到HSV模型的转换公式:

$$\begin{aligned} \max &= \max(R, G, B), \\ \min &= \min(R, G, B), \\ \delta &= \max - \min, \\ V &= \max, \\ S &= (\max - \min) / \max, \\ \text{if } S &= 0, H \text{ is undefined, else} \\ \text{if } R &= \max, H = (G - B) / \delta, \\ \text{if } G &= \max, H = 2 + (B - R) / \delta, \\ \text{if } B &= \max, H = 4 + (R - G) / \delta, \\ H &= H \cdot 60, \text{ if } H < 0, H = H + 360. \end{aligned} \quad (1)$$

式(2)是RGB模型到YES模型的转换公式:

$$\begin{bmatrix} Y \\ E \\ S \end{bmatrix} = \begin{bmatrix} 0.253 & 0.684 & 0.063 \\ 0.500 & -0.500 & 0.000 \\ 0.250 & 0.250 & -0.500 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \quad (2)$$

2.2 人脸掩膜处理

由 20×20 的彩色窗口图像得到的颜色分量图像,在被送入神经网络进行分类判别之前,必须先经过一个自动掩膜处理过程,用一个经由实验确定的 20×20 的人脸掩膜进行过滤处理,滤除人脸区域外的背景成分(图1),最后得到的掩膜处理后的图像包含364个象素点。

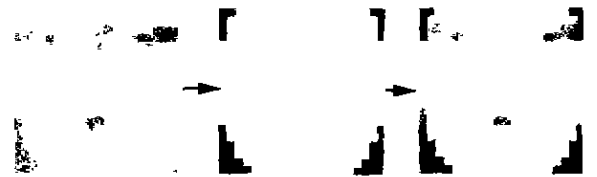


图1 人脸掩膜处理
Fig.1 Processing face mask

2.3 神经网络结构

本系统的每一级神经网络均采用基于反向传播(BP)算法的多层感知机(MLP)结构,其输入层节点对应于经过人脸掩膜处理的颜色分量图像,共364个节点,输出层采用单个节点,其值的1和0分别表示“是人脸”和“不是人脸”,而隐含层节点数则由经验确定为10。

2.4 神经网络的级联

多年来,人们提出了大量的模式识别方法,这些方法各有优点,但也不可避免地存在一定的局限性和适用条件。为了把各种方法有机地结合起来,取长补短,提高系统的整体性能,人们提出了多级神经网络的结构模型。多级神经网络结构的每一级子系统,通常都由特征提取模块和神经网络模块所组成。特征提取模块从输入数据中提取特征矢量,基于该特征矢量,神经网络模块在学习阶段对所有训练数据进行学习,而在工作阶段则对输入到本级的数据进行分类判别,包括识别输出和拒识输出两种情况,不妨假定:某一级的特征提取模块所提取的特征能有效地代表一部分输入样本的本质特征,因此这一部分样本在该级中将成功地得到识别,其余的样本则由于特征不清而被拒识,送入到下一级进行基于另外一种特征提取和神经网络结构的分类判别。

一个好的多级神经网络结构应满足两个条件:

(1) 有效设计每一级的特征提取模块, 尽可能使每一级的特征提取方法相互独立, 互为补充; (2) 合理设计各级神经网络模块, 使得该神经网络能有效地拒识那些在该级特征下模糊不清的样本, 从而使误识率尽可能降低。多级神经网络结构的整体性能在很大程度上依赖于每一级子系统误识率的高低, 而前几级的误识率尤其关键。

本系统以不同的颜色分量图像作为每一级神经网络的输入, 相当于为每一级神经网络分别设计了相互独立的特征提取方法。另一方面, 为了提高每一级神经网络的分类正确率, 降低误识率, 我们把 0.8 和 0.2 作为分类判别的上下限, 当网络输出大于 0.8 时判为“人脸”, 小于 0.2 时判为“不是人脸”, 而对那些介于 [0.2, 0.8] 之间的网络输出则拒识, 留待基于其它颜色分量的下一级神经网络作进一步分类。

3 实验结果与分析

3.1 训练/测试样本的选取

为了对系统进行计算机模拟实验, 我们从 Internet 上下载了 172 幅彩色图像, 并在此基础上用手工方法切割出 200 幅单张人脸图像和 200 幅非人脸图像, 把它们任意划分成两组, 各含 100 幅人脸和 100 幅非人脸图像, 分别构成训练样本集和测试样本集。为了保持原图像中目标的长宽比例, 我们采用正方形来切割人脸与非人脸目标, 确保不破坏目标的真实比例。

训练和测试样本集中的所有人脸样本图像, 均包含眼睛、鼻子、嘴巴这 3 个主要的人脸器官, 但基本上不包括头发、耳朵等部分, 以减小不同的人脸外观 (如有无头发、耳朵是否被遮挡等) 对人脸检测结果的影响, 另外, 样本图像中的人脸主要是白种人, 也包括少

量的黄种人, 其它人种目前暂不予考虑。而非人脸样本图像则既包括天空、森林、草地、房屋、室内布置等常见的背景图案, 也包括一些部分缺损的人脸图像, 以利于提高系统的分辨能力。

3.2 单级神经网络分类实验

首先通过彩色模型转换, 把训练集中的所有样本图像转换成不同的颜色分量图像, 以此作为神经网络的输入, 然后在此基础上对神经网络进行分类训练, 最后用测试样本对训练好的神经网络进行分类测试。

实验结果如表 1 所示, 表中的 RMS 误差即均方根 (Root Mean Square) 误差, 分类判别正确率、拒识率和误识率是用测试样本对训练好的神经网络进行分类测试时, 网络输出大于 0.8、介于 [0.2, 0.8] 之间以及小于 0.2 的测试样本在测试样本总数中所占的比率。

表 1 的实验结果具有以下两个特点: (1) 色度分量, 包括 HSV 模型的 H、S 分量以及 YES 模型的 E、S 分量, 其网络训练误差收敛缓慢, 当用测试样本对训练好的神经网络进行测试时, 这些色度分量的 RMS 误差普遍较大 (>0.2), 分类判别正确率也较低 ($<70\%$), 因此, 这些色度分量难以直接用于基于神经网络的人脸判别任务; (2) 除此之外的其它颜色分量, 包括亮度分量 V、Y 和 RGB 色度分量, 其网络训练速度相对较快, 测试样本集的 RMS 误差也大大减小, 对测试样本的分类判别正确率接近于实用要求。

由此可见, 由于只考虑图像的颜色成分, 而对人脸中眼睛、鼻子、嘴巴等器官的相对位置分布不加考虑, 色度分量的分类判别效果不佳, 而亮度分量 (以及与亮度高度相关的 RGB 分量) 代表图像的亮度特征, 包含了人脸的结构信息, 因此网络训练效果较好, 分类判别能力较强, 可直接用于基于神经网络的人脸判别过程。

表 1 单级神经网络分类结果

Table 1 Classified results for single-stage neural network

彩色模型	颜色分量	迭代次数	训练样本集 RMS 误差	测试样本集 RMS 误差	对测试样本作分类判别		
					正确率 (%)	拒识率 (%)	误识率 (%)
RGB	R	643	0.0031	0.1049	93	6.5	0.5
	G	797	0.0031	0.1366	89	10.5	0.5
	B	2289	0.0027	0.1965	78	21	1
HSV	H	3918	0.0165	0.3599	45.5	46.5	8
	S	3919	0.0233	0.3839	46.5	41	12.5
	V	765	0.0026	0.1083	89.5	10.5	0
YES	Y	1759	0.0014	0.1273	90	9.5	0.5
	E	4159	0.0059	0.2967	59	34.5	6.5
	S	1698	0.0018	0.2224	69.5	27	3.5

3.3 级联神经网络分类实验

尽管使用单级神经网络已经可以达到接近于实用的分类判别正确率,但如果要进一步提高则相当困难,一种行之有效的方法是把基于不同颜色分量的神经网络进行级联。

由前文可知,为了提高级联神经网络的整体性能,在前几级神经网络中,应该尽可能选用误识率低的颜色分量。考虑到 HSV 模型的 V 分量的分类判别误识率最低(0),本实验选用 V 分量作为第一级神经网络的颜色分量,然后在此基础上对第一级的所有拒识样本用其它颜色分量进行第二级分类判别,而第三级的分类判别则建立在第二级的拒识样本基础上,实验结果如表 2 所示,表中数据代表正确率、拒识率、误识率的百分比,而括号中的数据则表示相对于前一级(黑体字表示)神经网络增加或减少了多少个百分点。

由级联神经网络分类实验结果可知:(1)第二级神经网络的分类判别正确率普遍高于第一级,提高的幅度从 2 到 6 个百分点不等,说明通过神经网络的级联确实能够有效提高系统性能;(2)在第二级神经网络中,YES 模型的 E 分量的分类判别正确率提高了 6 个百分点,而误识率却只增加了半个百分点,其实验效果最好,因此,本实验选用 E 分量作为第二级神经网络的颜色分量,并在此基础上进行第三级神经网络分类实验;(3)在第三级神经网络中,不管是分类判别正确率还是误识率,其上升幅度均有所下降,最大仅 2

个百分点,说明随着神经网络级数的增加,系统性能的改善幅度也逐步下降。

如前文所述,要通过多级神经网络的级联来提高系统性能,关键的一点是尽可能使每一级的特征提取方法相互独立,互为补充,而在不同的颜色分量中,亮度分量与色度分量的差别很大,可视作是这种彼此独立的特征。基于亮度分量 V 和色度分量 E 的神经网络的级联,恰恰满足了这个要求,因此可以大大提高系统的总体性能。

由以上分析可知,尽管通过级联更多级神经网络可以进一步提高分类判别正确率,但实验结果表明,神经网络分类性能的改善在第二级最为显著,而在第三级已经大大减小,因此完全没有必要级联过多的神经网络,一般应以二或三级为限。

3.4 系统实验

基于以上分析,我们在整个人脸检测系统中采用二级神经网络级联方式,以 HSV 模型的 V 分量作为第一级神经网络的颜色分量,第二级神经网络则基于 YES 模型的 E 分量。

图 2 是用本系统对彩色图像进行人脸检测的几个实例,图中的白色小方框表示系统检测出的人脸区域。

4 结论

本文提出一种基于颜色信息与多级神经网络的人脸检测方法,首先通过对整幅图像进行尺度变化和窗

表 2 级联神经网络分类结果
Table 2 Classified results for cascading neural networks

网络级联	彩色模型	颜色分量	对测试样本作分类判别		
			正确率(%)	拒识率(%)	误识率(%)
第 1 级	HSV	V	89.5	10.5	0
		R	94.5(+5)	5(-5.5)	0.5(+0.5)
		G	92.5(+3)	7(-3.5)	0.5(+0.5)
第 2 级	RGB	B	93(+3.5)	7(-3.5)	0(+0)
		H	94.5(+5)	5(-5.5)	0.5(+0.5)
		S	91.5(-2)	5.5(-5)	3(+3)
	YES	Y	94(+4.5)	5.5(-5)	0.5(+0.5)
		E	95.5(+6)	4(-6.5)	0.5(+0.5)
		S	94(+4.5)	4.5(-6)	1.5(+1.5)
第 2 级	YES	E	95.5	4	0.5
		R	97(+1.5)	2(-2)	1(+0.5)
		G	97(+1.5)	2.5(-1.5)	0.5(+0)
第 3 级	RGB	B	97(+1.5)	2.5(-1.5)	0.5(+0)
		H	97.5(+2)	2(-2)	0.5(+0)
		S	95.5(+0)	2(-2)	2.5(+2)
	YES	Y	97(+1.5)	2.5(-1.5)	0.5(+0)
		E	96.5(+1)	2(-2)	1.5(+1)
		S	96.5(+1)	2(-2)	1.5(+1)



图 2 本系统的一些人脸检测结果

Fig. 2 Some face detection results by the present system

口扫描,解决待测人脸的位置、大小问题,然后通过基于不同颜色分量的神经网络的级联,对扫描得到的彩色窗口图像进行人脸分类判别,最后将检测到的所有人脸的位置和大小信息输出。

本方法采用彩色图像的亮度分量和色度分量作为级联神经网络的分类特征,从而避免了既复杂费时、效果又不一定理想的特征抽取过程,同时又充分利用了彩色图像中丰富的颜色信息,有效提高了人脸检测的正确率,加快了人脸检测速度。

REFERENCES

- [1] Kwon Y H, Lobo N V. Face detection using templates, *IEEE 12th Int. Conf. on Pattern Recognition*, 1994, **1**: 761-767
- [2] Yuille A, Cohen D, Hallinan P. Feature extraction from faces using deformable templates, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1989: 104-109
- [3] Pentland A, Moghaddam B, Starner T, et al. View based and modular eigenspace for face recognition, *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, 1994: 84-91
- [4] Rowley H A, Baluja S, Kanade T. Neural network-based face detection, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1998, **20**(1): 23-38
- [5] Dal Y, Nakano Y. Face-texture model based on SGLD and its application in face detection in a color scene, *Pattern Recognition*, 1996, **29**(6): 1007-1017
- [6] Lee C H, Kim J S, Park K H. Automatic human face location in a complex background using motion and color information, *Pattern Recognition*, 1996, **29**(11): 1877-1889
- [7] Roberts A, Ford A. Colour (color) space conversions, http://www.zenger.informatik.tu-muenchen.de/~gdy/info/csc/COL_.htm, 1996