

双向层级结构分类器及其在面部特征点定位中的应用^①牛志恒^{②*} 山世光^{③**} 陈熙霖^{**}

(* 哈尔滨工业大学计算机科学与技术学院 哈尔滨 150001)

(** 中国科学院计算技术研究所 中国科学院智能信息处理重点实验室 北京 100190)

摘 要 设计了一种新颖的双向层级结构分类器,并将其应用于面部特征点的定位。该分类器在训练过程中对正例样本和反例样本交替进行重采样,因而与采用传统的单向层级分类器的面部特征定位方法相比,采用这种新的双向层级结构分类器的方法具有以下优点:可以应对大规模的数据集;可以处理存在复杂变化的正反例样本;而且无论是在训练过程还是在测试过程,其算法都能快速地过滤大量的“易分”样本,执行效率非常高。在两个公开测试数据库上的实验结果表明,采用双向层级结构分类器的方法可以实现准确、快速的特征点定位。

关键词 双向层级结构, 分类器设计, 正(反)例样本, 面部特征点定位

0 引言

在大多数模式识别和计算机视觉任务中,对目标的检测和定位^[1]是十分重要的。近些年许多国内外研究者致力于目标检测和定位问题的研究并且取得了丰硕的研究成果。但是到目前为止,这个问题仍然没有得到有效的解决,尤其是在非理想成像条件情况下的检测和定位问题。现有的目标检测和定位方法基本上可以归为三类:基于底层特征和先验规则的方法,基于目标模板和变形模型的方法,以及基于高层特征(高维统计量)和分类器的方法。前两种方法在数据规模较小、实验条件较理想的条件下有良好的效果,但是在数据规模较大、数据的分布和变化也较为复杂的情况下,检测和定位就会出现较大的误差。而第三种方法由于提取了稳定的特征再加上分类器具有强大的判别能力,因而有较高的检测或定位精度和较好的抗干扰能力,近年来得到了较为广泛的使用。对于这一类方法,在特征层面,常用的有 Gabor 特征、Haar 特征等,在分类器层面常见的有神经网络、支持向量机(SVM)、Boosting 等。其中 Haar 特征的积分图算法和 Boosting 的层级结构,使得特征提取和分类过程非常高效,于是 Haar 特征和 Boosting 分类器的组合成了目前最流行的检测或

定位方法。本文提出的方法属于这种先提取特征进而进行分类的方法,但是与传统的方法不同,这种方法在分类器结构设计上采用了双向的层级结构,从而使得该算法无论在精度上还是在速度上都比传统的方法有所提高。

1 传统层级分类器

层级结构的分类器^[2]已被广泛用于目标检测和模式分类。图 1 是这种层级分类器的结构和流程。输入的待检测窗口依次经过层级结构的子分类器,当子分类器判断为反例时就直接将该目标判定为反例,而如果子分类器判定为正例时则要经过下一个分类器,直到最后一个子分类器输出为正例时才判定该目标为正例。因此大量的反例样本可以快速地在较早的阶段就被抛掉,从而提高了算法的效率。

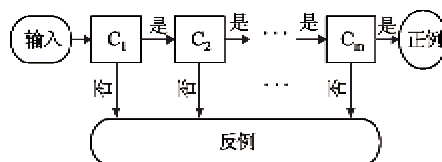


图 1 传统单向层级分类器的结构

① 863 计划(2007AA01Z163)和自然科学基金(U0835005)资助项目。

② 男,1980 年生,博士生;研究方向:计算机视觉,模式识别;E-mail: zhniu@jdl.ac.cn

③ 通讯作者, E-mail: sqshan@jdl.ac.cn

(收稿日期:2009-04-01)

但是这种传统的层级分类器在训练时只是在反例样本上通过自举(bootstrap)进行重采样,并没有考虑正例样本的重采样。而且实际上对于很多问题,正例样本的规模也是非常大的。加大正例样本的规模,可更加充分涵盖在实际情况中所能出现的样本,有助于提高分类器的性能,增强算法的推广能力。传统的单向层级分类器对于正例样本是作为一个整体考虑的,训练时使用所有的样本。一旦正例样本规模过大,便只能采用随机采样的策略选取一个较小的子集用于训练。

传统的单向层级分类器难以应对大规模的正例样本的另一个原因在于过于庞大的样本数据往往对应着复杂的模式变化,这就使得样本间的散度变大,从而导致训练过程难以收敛,计算代价也随之增大。这些样本由于采集环境和条件的不同,对应的模式极其复杂。因此,传统的层级分类器对这种大规模正例样本集合的训练往往很难奏效。

于是我们希望在正例样本上也通过自举进行重采样,把整个分类器分解成一些子分类器。按照这种思路,子分类器在进行正例判断时,也可以快速确认,而不必经过所有分类器的计算才得到最终的结果。这是因为对于简单的正例样本只用很少的特征和分类器就可以准确快速地确认,而对于那些情况复杂的正例样本则需要更多的特征、更复杂的分类器耗用更多的时间来分析。这样就避免了传统单方向层级分类器必须要通过最后一个子分类器才能判断是否正例的问题。

2 双向层级结构分类器

2.1 分类器结构设计

针对上节所述的传统层级分类器对于大规模正例样本集合训练过程存在的限制,本文提出了一种新的双向层级结构的分类器,算法思想如图2所示。所谓的双向即在水平方向通过自举对正例样本进行重采样,同时在竖直方向通过自举对反例样本进行重采样。图中的每一个小虚线框代表了一个子分类器,水平和竖直两个方向是两个级别的关系。每一个虚线框包含的弱分类器构成的整体形成一个水平子分类器,我们称之为第一级。每个水平子分类器包含的内部的若干二级子分类器称为竖直子分类器。

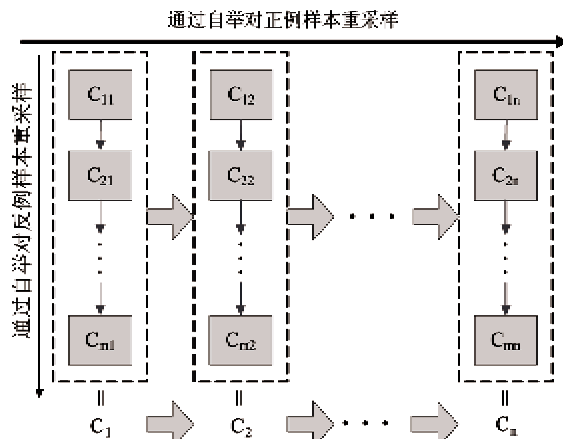


图2 双向层级分类器的结构

这样的双向层级分类器的设计有两个好处：一是在训练时可以快速地把正反例样本分开批次，每个子分类器专门应对一部分样本，也就是说把整个的训练集分而治之，加快算法收敛，这个分批训练的过程是通过在样本集上自举重采样自动完成的；二是在测试时如果输入数据比较简单，则使用简单的分类器就可以快速完成分类，进而达到加速算法的目的。

2.2 分类器训练过程

传统的层级分类器训练时在反例样本上通过自举进行重采样,为了保证最后的正例样本的检测率,一般每个子分类器设定很高的正例检测率(一般在99%以上),而反例样本的检测率只要维持在一个较低的水平即可(如50%左右)。这是因为层级分类器在逐级分类时一旦判断为反例就会把样本抛掉,立即判定为反例。所以最后能通过全部的子分类器的反例样本就会非常少,也就是可以保证较低的误检率。而对于双向层级结构的分类器,不需要前几级的水平子分类器有很高的正例检测率,可以像在传统层级分类器里设定反例检测率一样设定一个宽松的值。这样在水平方向前几级的子分类器会比较轻巧,可以快速地过滤大量的正例样本。

如图3所示,我们先在所有的正例样本随机选取一定数量的样本,然后用这个样本子集和所有的反例样本进行传统的单向层级分类器的训练,训练得到水平方向的子分类器。然后通过自举的方式对所有的正例样本进过滤,将那些没有被正确分类的样本再重新随机采样用于下一个水平子分类器的训练。其中每一个水平子分类器的训练过程也是通过自举不断重采样的过程,其训练流程也是如同图3,不同之处在于,竖直子分类器自举重采样的过程是

在反例样本上进行的,而每一个水平子分类器包含了一系列的竖直分类器。

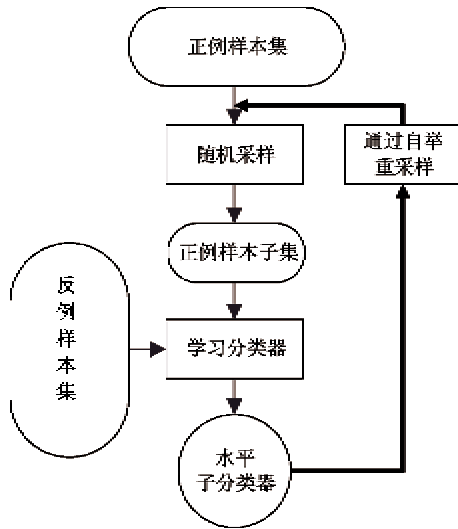


图3 水平分类器的训练流程

2.3 分类器决策过程

给定待测图像,我们在测试图像的每个位置上使用训练得到的分类器对其进行判定。每个水平子分类器通过对这些位置上的滑动窗口进行分类得到一系列的目标窗口,而对于定位而言,需要给定一个目标的位置,我们用如下方式来计算第 i 个水平子分类器 C_i 的定位结果:

$$L_i = \arg \max_{1 \leq p \leq N_i} \sum_{q=1}^{N_i} \varphi(p, q) \quad (1)$$

其中假定子分类器 C_i 检测到的窗口数目是 N_i , $\varphi(p, q)$ 是一个计数函数,如果 p, q 两个窗口中心位置的间距小于一定距离,则 $\varphi(p, q)$ 为 1,否则为 0。这样水平子分类器定位的位置就是第 L_i 个检测窗口的中心。

最终我们要综合所有子分类器的结果。本文给出了两种决策的方法,一种是对多个水平子分类器进行加权求取平均位置,另一种是按照水平方向的次序报告检测的位置,一旦检测到了目标的位置则停止后面子分类器对图像的分析。之所以可以采用第二种方式,是因为定位问题和检测问题有很大的区别,检测时不知道目标是否存在,也不知道存在多少个目标,而在定位问题里确定有且只有一个目标。所以在检测正例时也可以采用层级的方式来定位,对于简单的图像,目标易于检测和区分的图像,那么在前面的子分类器找到目标之后,后面的子分类器就没有必要再检测了,从而达到加速算法的目的。

2.4 双向层级分类器在面部特征点定位中的应用

本文将双向层及分类器应用到面部特征点定位问题中,进行双眼、鼻尖和两个嘴角这 5 个面部关键特征点的定位。下面我们以眼睛为例进行叙述,其它的特征点定位方法相同。首先在图像上提取高维的特征,然后训练分类器来区分眼睛区域和非眼睛区域,以提高眼睛定位的精度。在训练过程中,我们以眼睛为中心截取的矩形图像作为正例样本,以偏离眼睛中心一定距离以外截取的矩形图像作为反例样本,如图 4 所示。在定位时需要前端的人脸检测器给出一个大致的人脸位置,然后在人脸特定的区域通过双向层级分类器进行眼睛的检测,最后根据分类器检测的结果通过上述的决策方式输出眼睛的位置。

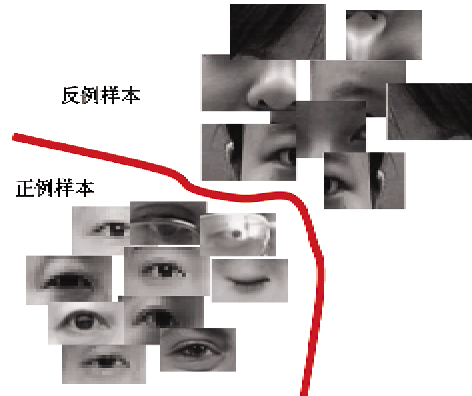


图4 用于眼睛定位的训练集合示例

3 实验

3.1 实验方案

本文采用了多个采集环境各异,姿态、表情、光照等变化复杂的人脸数据库。并且为了体现算法优异的推广性能,训练和测试是在不同的人脸库上进行的。用来做训练的人脸库包括 FERET^[3]、BANCA^[4]、CAS-PEAL^[5] 等数据库,一共 30000 多张人脸图像。用来做测试的人脸库包括 XM2VTS^[6] (共 1180 幅图像)、BioID^[7],共 1521 幅图像。

为了使得样本空间更加稠密,我们对正例样本通过适当的旋转和缩放进行了扩充,扩充后的正例样本集达到 150000。所以我们的数据规模是十分庞大的,这对于传统的分类器是很难处理的。

根据 2.2 节所述的流程,我们先从这 150000 正例样本里随机选取 10000 样本和所有的反例样本进行传统的单向层级分类器的训练,得到子分类器

C_1 。本文设定一级子分类器的检测率为 50%,通过自举的方式,在 150000 正例样本中,大约有 75000 样本没有被 C_1 检测到,然后再从其中随机选取 10000 样本和所有的反例样本训练得到 C_2 ,经过几级之后总体的强分类器会达到很高的检测率和极低的误检率。在这个过程中子分类器不断加强,第一个子分类器只包含了几十个弱特征,而最后的一个子分类器则需要几千甚至更多的弱特征。

在图 2 中的子分类器从上到下,从左到右分类器是由简到繁的,子分类器越来越复杂,体积越来越庞大,具体体现就是构建子分类器所需的弱特征越来越多。表 1 给出了双向层级结构的分类器中的子分类器沿两个方向的变化趋势。其中 $\uparrow$ 表示越来越大, $\downarrow$ 表示越来越小。在竖直方向上,子分类器处理的样本越来越复杂,样本的规模越来越大,训练所耗费的时间也越来越多。由于是在反例上进行自举,所以误检率下降得很快(指数级别),检测率也稍有下降;在水平方向上,同样是子分类器处理的样本越来越复杂,样本的规模越来越大,训练所耗费的时间也越来越多。由于是在正例样本的自举过程,所以检测率很快地提高了,随之误检率也会稍有上升(线性级别)。

表 1 子分类器的属性和趋势

	样本复杂度	样本规模	检测率	误检率	耗时
竖直方向 $\uparrow$ (反例样本)	$\uparrow$	$\uparrow$	$\downarrow$	$\downarrow$	$\uparrow$
水平方向 $\uparrow$ (正例样本)	$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$

3.2 算法性能评价准则

评价定位精度最简单的准则是求取在所有待测图像上自动定位和手工标定的平均误差。考虑到按像素度量的误差与图像中人脸的尺寸有很大关系,所以有必要进行归一化的处理。为更好地描述误差的具体分布,文献[8]首先提出以式

$$Err = \frac{\max(d_l, d_r)}{\|C_l - C_r\|} \quad (2)$$

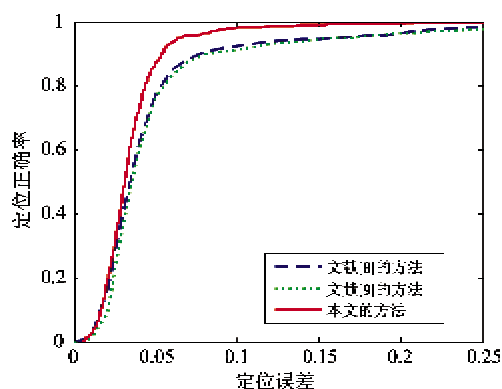
度量眼睛定位的误差,其中 d_l, d_r 分别是与手工标注位置相比两只眼睛自动定位的误差, C_l, C_r 分别是手工标注的眼睛位置。之后这种方法被学者普遍使用。对于一幅人脸图像,用两眼自动定位和手工标定的最大相对误差来代表这幅图像上的定位误差,相对误差是指用这幅图像手工标定的两眼间距进行归一化后的误差。然后在所有的图像上统计小于特定误差的图像比率,从而得到一条累积曲线 y

$= P(Err \leq x)$, 从这条曲线上我们可以直观地看出定位误差的分布情况。

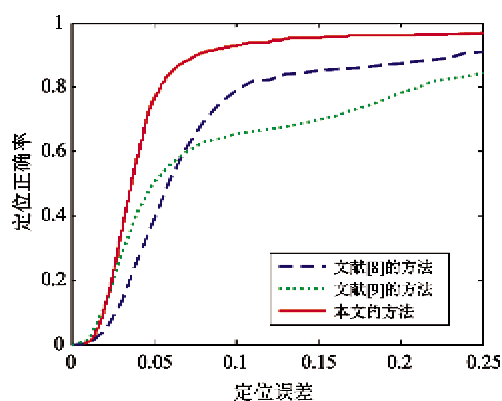
3.3 实验结果分析

3.3.1 与其他方法比较

文献[8]和文献[9]都在 XM2VTS 和 BioID 两个库上按照 3.2 节的评价准则报告了眼睛定位的实验结果。本文方法与这两个方法的对比由图 5 给出。可以看到,在这两个人脸库上,本文方法的精度都是最高的,尤其是在 BioID 库上本文方法的精度要明显优于文献[8,9]的方法。XM2VTS 人脸库的图像背景单一,很少出现复杂的人脸变化,而 BioID 人脸库图像背景较为复杂,人脸变化比较复杂,这也表明本文方法在复杂背景下也很鲁棒。



(a) XM2VTS 人脸库



(b) BioID 人脸库

图 5 本文方法与前人方法的双眼定位精度比较

3.3.2 不同特征点的比较

与眼睛的定位相比,在对鼻尖和嘴角的实验中,定位的性能曲线如图 6 所示。在 XM2VTS 数据库上由于姿态光照比较统一,总体的定位精度要高于在 BioID 上的结果,尤其是鼻尖的定位在 XM2VTS 上的结果要明显高于在 BioID 上的结果。我们分析其中原因,认为主要是鼻尖最明确的特征是正面的高光,

所以在 XM2VTS 的精确正面光照情况下,鼻尖的高光特征明确且稳定,而在 BioID 人脸库上不确定的

环境光对鼻尖的定位产生了较大的干扰。

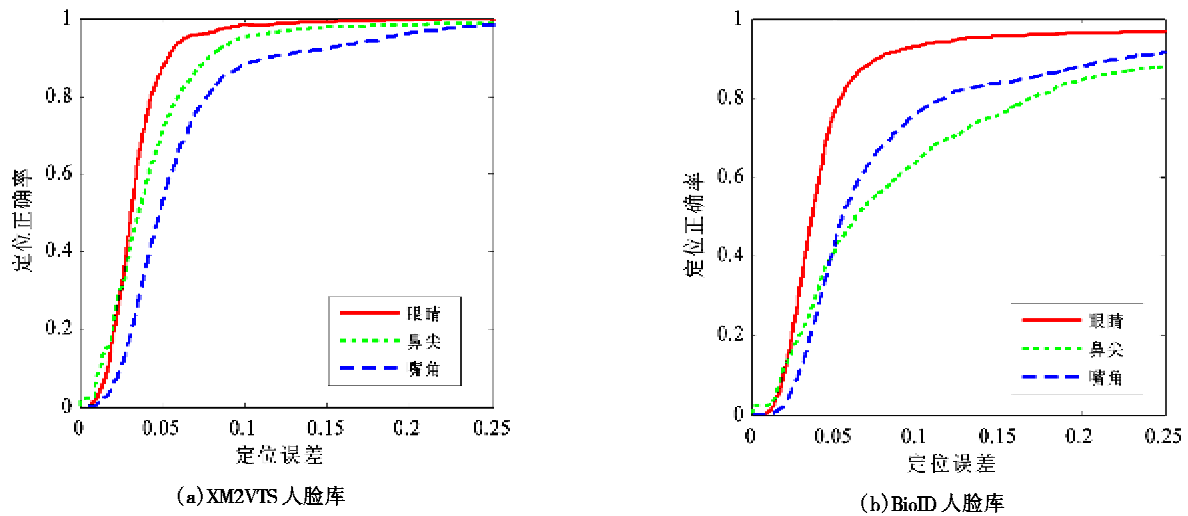


图 6 不同面部关键特征点之间的定位精度比较

通过上面的实验结果可以看到,在面部的几个关键的特征点中,相对来说眼睛还是比较稳定的特征点,所以定位的精度最高。鼻尖和嘴角相对于眼睛提供的结构信息比较少,定位精度较低,这在特征点人工标注的过程中也能看出来,标注鼻尖和嘴角

的歧义性比较大,即不同的人对鼻尖和嘴角的位置有不同的主观定义,而对于眼睛位置的判断则比较一致。最后在两个数据库里的部分人脸图像上的定位结果由图 7 给出。



图 7 人脸库中部分图像的面部特征点定位结果

3.3.3 性能分析

根据 2.3 节所述,本文验证了 4 种不同定位策略。双向层级结构的分类器可以使用全部的正例样本,但是对于单向结构的层级分类器,只能通过随机采样使用其中的一部分。在 150000 正例样本中,我们通过随机抽取其中的 10000 正例样本构建了传统的单向层级结构分类器,并且同样训练了 3 个独立的单向结构分类器进行融合,具体结果如表 2 所示。在表 2 中,定位的精度是在 $Err < 0.1$ 的条件下取得的,平均每幅图像定位的耗时是在 2.4GHz Intel 奔腾

CPU 的普通台式机上测得的结果。从中我们可以看到,对所有的水平子分类器进行加权的精度是最好的,但是比较耗时,而采用双向层级结构的分类器在精度依然较高的情况下,大大地降低了定位所消耗的时间,可以满足实时的应用环境。传统的单向结构层级分类器在训练时难于收敛,构成强分类器所需的弱特征很多,所以在应用时比较耗时,而且由于每次训练出来的强分类器性能较为相似,所以通过融合的策略精度提高有限,而定位消耗的时间则成倍地增加。

表2 不同方法的定位精度和计算速度

	双向结构 (加权)	双向结构 (层级)	单向结构 (融合)	单向结构 (随机采样)
XM2VTS	98.1%	97.0%	97.2%	96.8%
BioID	93.0%	91.4%	91.1%	91.0%
平均耗时	~ 1200ms	~ 25ms	~ 1300ms	~ 420ms

4 结 论

本文提出了一种基于双向层级结构的分类器并将其应用于面部特征定位中。其算法在训练过程中对正例样本和反例样本交替进行重采样,从而使得无论是在训练还是在应用过程中,都能快速地抛弃大量的“易分”样本,因而执行效率非常高。这种分类器的设计克服了采用传统的算法无法有效利用全部大规模的正例样本,以及算法难于收敛的问题。总体上说,采用双向结构学习的算法自动将一个大规模的样本集分而治之,从而在总体上降低了算法的复杂度,使得算法能够较快地收敛。这使得我们可以通过增加样本的规模 and 变化,提高算法的精度,并可通过分级的结构提高算法的效率。

参考文献

- [1] 曾建超,柳慧群,林学. 基于规则的物体识别定位方法. 高技术通讯,1995,5(4):21-24
- [2] Viola P, Jones M. Robust real-time face detection. *International Journal of Computer Vision*, 2004, 57(2): 137-154
- [3] Phillips P, Wechsler H, Huang J, et al. The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 1998, 16(5): 295-306
- [4] Messer K, Kittler J, Sadeghi M, et al. Face authentication test on the BANCA database. In: *Proceedings of the International Conference on Pattern Recognition*, Cambridge, UK, 2004. 523-532
- [5] 张晓华, 山世光, 曹波等. CAS-PEAL 大规模中国人脸图像数据库及其基本评测介绍. 计算机辅助设计与图形学学报, 2005, 17(1): 9-17
- [6] Messer K, Matas J, Kittler J, et al. XM2VTSDB: The extended M2VTS database. In: *Proceedings of International Conference on Audio and Video-based Biometric Person Authentication*, Guildford, UK, 2003. 72-77
- [7] Jesorsky O, Kirchberg K. BioID face database. <http://www.bioid.com/downloads/index.php>, 2001
- [8] Jesorsky O, Kirchberg J, Frischholz W. Robust face detection using the hausdorff distance. In: *Proceedings of the International Conference on Audio and Video-based Biometric Person Authentication*, Halmstad, Sweden, 2001. 90-95
- [9] Hamouz M, Kittler J, Kamarainen J. et al. Affine-invariant face detection and localization using gmm-based feature detector and enhanced appearance model. In: *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, Seoul, Korea, 2004. 67-72

A novel bidirectional cascaded classifier and its application in facial feature localization

Niu Zhiheng*, Shan Shiguang**, Chen Xilin**

(* School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001)

(** Key Lab of Intelligent Information Processing, Chinese Academy of Sciences (CAS)

Institute of Computing Technology, CAS, Beijing 100190)

Abstract

A novel bidirectional cascaded classifier was designed and it was applied to facial feature localization. In the training phase, the classifier resamples the positive and negative samples alternately, so compared with the traditional facial feature localization method based on a unidirectional cascaded classifier, the method based on the new classifier has the following advantages: it can cope with large scale data sets, can deal with complicated variations of the positive and negative training samples, and either in the training stage or in the test process its algorithm can rapidly filter large number of simple samples, thus obtaining a very high efficiency. The results of the extensive experiment on two public face databases verified the effectiveness and efficiency of the method based on the new classifier.

Key words: bidirectional cascaded architecture, classifier design, positive/negative samples, facial feature localization