

文章编号: 1007-1482 (2011) 01-0038-0043

• 图像技术 •

## 基于面部动作单元组合特征的表情识别

欧阳琰<sup>1</sup>, 桑 农<sup>2</sup>

(华中科技大学 图像识别与人工智能研究所, 武汉 430074)

**摘 要:** 人脸表情可以被看作是由面部表情编码系统 (FACS) 定义的不同面部运动单元的组合。不同于人脸图像的灰度、纹理等表象特征, 基于面部运动单元的表情混合特征能够更准确地描述表情, 然而, 面部运动单元很难精确定位, 为了避免这个问题, 在前人的工作中通过将图像分成许多子块, 并从子块中提取面部运动单元信息来组成基于面部运动单元的表情成分特征。在此基础上, 本文首先通过对人脸图像的眼睛和口部作粗定位, 接着根据眼睛和口部的水平位置, 提取眼睛区域、口部区域和鼻子区域的图像子块, 然后对每个子块提取 Haar 特征, 并采用错误率最小策略从这些子块中选出面部运动单元组合特征, 最后使用组合特征进行学习得出弱分类器, 并嵌入到 Boost 学习结构中构造出强分类器。通过在 Cohn-Kanada 数据库上的测试, 证明本文的方法能够取得很好的表情分类效果。

**关键词:** 面部表情识别; 面部运动单元; 特征组合; boost 学习

**中图分类号:** TP391.44

**文献标识码:** A

## Facial expression identification based on combinational feature of facial action units

OUYANG Yan<sup>1</sup>, SANG Nong<sup>2</sup>

(Institute for Pattern Recognition &amp; Artificial Intelligence, Huazhong University of Science and Technology, Wuhan 430074, China)

**Abstract:** Facial expressions may be described as combination of facial action units defined by facial action coding system. Unlike appearance features of face images, such as gray and texture, the combination feature of facial action units can describe the facial expressions better. However, it is difficult to detect facial action units accurately. So, many previous works try to divided face image into local patches, and extract the information of facial action units to compose the compositional features of facial expressions. According to these works, in this paper we firstly locate the position of eye and mouth in face images, and then divide face images into local patches due to the position of eye and mouth, after that extracted Haar features from each patches and use a minimum error based combination strategy to build combinational feature of facial action units from these features of patches, then use combinational feature to build weak learners. Finally boosting learning structure is used to build the final strong learner. In the experiment on Cohn-Kanada database, the method described in this paper has a promising performance.

**Key words:** facial expression identification; facial action units; feature combination; boost learning

收稿日期: 2010-12-23

作者简介: 欧阳琰 (1984-), 男, 在读博士。研究方向: 人脸表情识别。

0 引言

早在 1978 年美国心理学家 Ekman Paul 和 Friesen 就开发了面部动作编码系统( FACS) [1], 该系统通过研究脸部肌肉群的运动, 描述了不同的面部动作。虽然面部动作单元( AU) 能够非常精确的表述面部表情, 但是由于很难精确定位, 因此很少有人采用它来识别人脸表情。先前文献中的表情识别方法主要分为 2 个步骤: 提取特征和机器学习, 提取特征的主要方法有: Gabor 特征提取方法、Haar 特征提取方法和 LBP 方法; 而机器学习的方法主要是: Adaboost 方法、SVM 方法和神经网络方法。例如: 文献 [2] 中采用了 LBP 特征提取和 SVM 方法对 Cohn - Kanade 数据库 [3] 上的表情图片进行了识别, 取得了很好的效果; 文献 [4] 通过比对指出首先通过 Adaboost 方法选取 Gabor 特征, 最后使用 SVM 进行表情分类的结果最好; 文献 [5] 采用 Adaboost 方法对 Gabor 特征进行选取, 并将选取的特征与 SVM 结合起来进行表情分类, 并在 JAFFE 数据库上得出了很好的结果。

人脸的每一种表情都可以用一些带有特殊空间信息的特征来表述。Gabor 特征、Haar 特征和面部动作单元( AU) 都是表情信息的一种描述形式, 但是他们也有不同。Gabor 特征、Haar 特征等主要是发掘表情图像中的空间信息, 而面部动作编码则是心理学上对表情信息的定义, 因此使用面部动作编码作为表情识别的特征将更有说服力。FACS 中一共定义了 44 种面部动作编码, 每一种表情都是由不同的面部动作编码组成 [6], 例如: 愤怒的表情是由 AU4( 额眉低垂)、AU5( 上眼睑上抬)、AU7( 眼睑紧凑)、AU15( 唇角下压) 和 AU24( 口唇压紧) 组成等等 [7]。

目前, 通过面部动作编码进行人脸表情识别已经成为表情识别方法中的热点之一, 国外已经有许多针对面部动作单元( AU) 定位和检测及其在表情识别中应用的研究: 文献 [8] 在模板匹配中结合人脸五官的上下文信息, 实现了面部动作单元的精确定位; 文献 [9] 以 AAM 标记出的人脸特征轮廓为基础, 采用基于分割的 SVM 方法对视频流中的面部运动单元进行检测; 文献 [10] 使用带有面部运动单元信息的成分特征来进行表情识别。本文在研究了文献 [10] 中的表情识别方法后, 提出了一种新的基于

面部运动单元的表情识别方法, 该方法首先对人眼与嘴巴进行初步定位, 使得面部运动单元的信息提取更准确, 同时特征的维数也更低, 接着采用新的特征组合方法和决策树来构建弱分类器, 最后使用 Adaboost 学习得到最终的强分类器并应用到表情识别上。下图为本文方法的整体流程图:

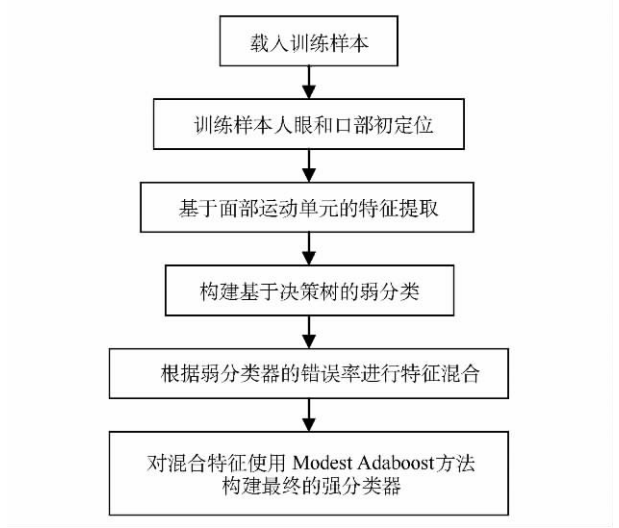


图 1 本文方法整体流程图

1 基于面部运动单元的特征提取

由于精确的提取面部运动单元是一件非常困难的事, 因此文献 [10] 中提出了一种局部分块的特征提取方法避开了定位面部运动单元的难题, 但是这种分块策略并没有考虑到人脸五官的初步定位, 提取特征的时候包含了一些的非面部运动单元的信息, 使得特征的维数上升, 计算量偏大。因此, 本文首先根据图像灰度信息采用灰度投影的方法 [6] 初步定位出眼睛、口部的大致位置, 最后根据这些初始位置使用类似于文献 [10] 中的特征提取策略提取面部运动单元信息。

1.1 人脸特征初步定位

文献 [10] 和文献 [11] 中训练和测试样本图像集是归一化后的 Kohn-Kanade 表情数据库中的表情图片( 如图 2( a) ), 因此, 本文也采取文献 [10] 和 [11] 中按照双眼间距对归一化后的图像集作为测试算法精度的数据( 如图 2( b) )。本文在对样本集提取表情特征之前, 对训练样本集中的人脸表情作了眼睛和口部水平线的初步定位, 用于提高识

别精度。

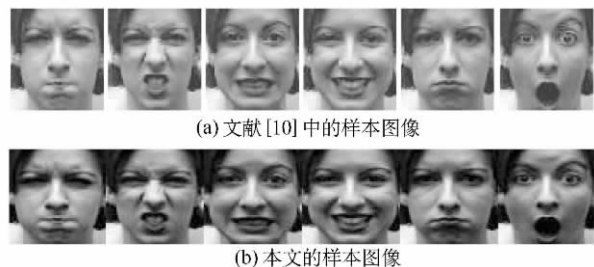


图2 样本图像

本文使用了文献[6]中的人眼和口部初步定位的方法。文献[6]根据人脸的五个器官之中眼睛部分的灰度梯度差最大,采用邻近点灰度差分公式计算该灰度梯度差,并设定阈值分离出眼睛区域。其算法流程<sup>[6]</sup>为:

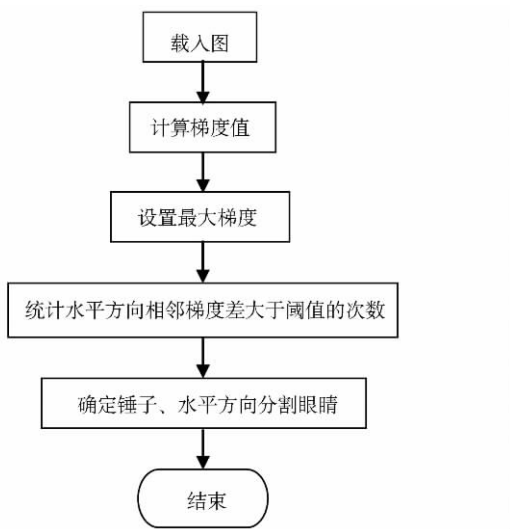


图3 人眼定位方法流程

由于是对归一化后的图像进行眼睛定位,因此,本文在上述方法之中加入了人脸的几何知识,从而获得极高的定位精度。在人眼的精确定位之后,本文接着使用文献[6]中关于口部水平中心线的定位方法,对口部进行定位,图4为定位后的图像。

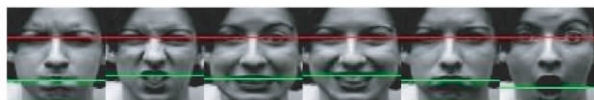


图4 眼睛和口部定位后的表情图

## 1.2 面部运动单元特征提取

如今采用面部运动单元作为表情识别的特征已经成为表情识别的热点,图5的第一行是 FACS 系

统中对人脸面部肌肉动作的标记,从该标记中我们可以发现,面部运动单元主要分为上半脸运动单元和下半脸运动单元,根据这个特点,结合1.1节中人眼和口部的初步定位,对面部表情图中局部面部运动单元的特征进行了提取。方法如下:

假定图像大小为  $m \times m$ ,本文首先以人眼水平中心线为水平中心,以  $m/8$  为子窗口水平移动步长,以  $m/8$  为垂直方向上的起始位置,提取大小为  $m/4 \times m/4$  的5个子窗口作为上半脸面部运动单元子块集  $\{P_i\}, i=1,2,3,4,5$ ,接着以口部水平中心线为水平中心,以  $m/8$  为子窗口水平移动步长,以  $m/8$  为垂直方向上的起始位置,提取大小为  $m/4 \times m/4$  的5个子窗口作为下半脸面部运动单元子块集  $\{Q_i\}, i=1,2,3,4,5$ ,最后以人眼与口部水平中心线的中点位置作为水平中心,以  $m/8$  为子窗口水平移动步长,以  $m/4$  为垂直方向上的起始位置,提取大小为  $m/4 \times m/4$  的3个子窗口作为人脸中部的面部运动单元子块集  $\{W_i\}, i=1,2,3$ ,这样就一共提取了13个子区域,具体提取示意图如图5:

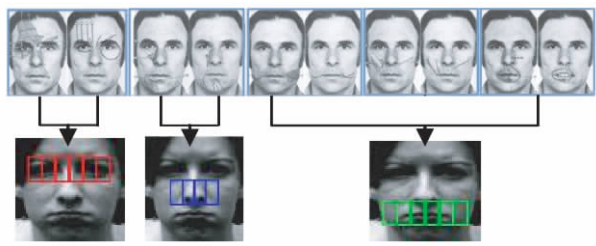


图5 局部子块与面部运动单元对应关系

图5中的第一行表示人脸运动单元的分布,第二行为使用本文提取区域的方法,提取的对应子块,可以看到基于人眼与口部初步定位的区域提取,准确的包含了绝大部分的运动单元。

在提取了运动单元区域之后,本文使用4种特征模板<sup>[12]</sup>来计算每个子块区域中 Haar 特征,子块集  $\{P_i\}$  中的 Haar 特征集合为  $\{\xi_i\}, i=1,2,3,4,5$ ,同理子块集  $\{Q_i\}$  的 Haar 特征集合为  $\{\psi_i\}, i=1,2,3,4,5$ ,子块集  $\{W_i\}$  的 Haar 特征集合为  $\{\zeta_i\}, i=1,2,3$ 。



图6 Haar 特征模板

2 基于决策树的弱分类器

不同于文献 [10]、[12] 中构造 Haar 特征弱分类器的方法, 本文使用决策树来构造弱分类器, 树分类器有两个优点: 1. 树所体现的语义信息容易直接用逻辑表达式表示出; 2. 由于只需要一系列的简单查询就可完成分类, 因此树形分类器的分类速度很快<sup>[13]</sup>。具体构造算法如下:

表 1 决策树结点生成算法

算法 1 决策树结点生成算法
1. 给定一组包含 $n$ 个样本的训练集 $S = (x^1, y^1), \dots, (x^m, y^m)$ , $x^j$ 表示样本的特征空间 $x^j = (x_i^j), i = 1, \dots, n, j = 1, \dots, m$ $y^j$ 表示每个样本的标记, 取值范围为 $\{-1, +1\}$
2. 找出 $x^j = (x_1^j, \dots, x_n^j)$ 中每一维特征的判决阈值, 使其区分 $S$ 的错误率最小;
3. 找到能够以最小错误率区分 $S$ 的第 $i$ 维特征, 并按下面的方法建立一个树的结点: a. 对第 $i$ 维特征使用判决式 $x_i > \theta$ 进行预测 b. 根据预测的正确与否, 生成结点的两条分支, 并指向不同的叶子

表 2 决策树生成算法

决策树建立算法
1. 使用结点生成算法构造根结点
2. 选择错误率最大的叶子
3. 接着构造与选中叶子相关联训练样本的子结点
4. 使用构造的子结点替代叶子
5. 重复步骤 2 和 4 直到所有叶子的错误率为 0, 或者达到了指定的查询次数

3 特征混合

人脸识别中提取成分特征的方法主要分为两种策略: (1) 首先将所有的图像特征放入一个成分特征池中, 然后使用 Adaboost 方法从特征池中提取最有效的成分特征, 并用这些成分特征来生成强分类器; (2) 在使用 AdaBoost 生成强分类器的过程中, 每次迭代选择一个成分特征, 在强分类器生成的同时, 成分特征也同时选取完毕<sup>[10]</sup>。本文使用的特征组合方法类似于第一种策略。根据表情的运动单元构

成信息<sup>[7]</sup>, 可以发现决定表情的人脸上半部运动单元主要为额眉处的肌肉运动、眼睑处的肌肉运动, 决定表情的下半部运动单元主要为口角处的肌肉运动、口唇的肌肉运动和唇角的肌肉运动。根据这些特性, 本文采用以下的特征组合策略。

表 3 特征组合算法

算法 2 特征组合算法
1. For $l = 1, \dots, n$ $n$ 表示可能特征组合的总个数
2. 从 $\{\xi_i\}$ 中随机选取 4 个子块特征集, 从 $\{\zeta_i\}$ 中随机抽取 1 个子块特征集, 从 $\{\psi_i\}$ 中随机抽取 3 个子块特征集, 将这 8 组特征组成新的混合特征 $\delta_l$ 。
3. 采用算法 1 中的弱分类器计算混合特征分类的错误率 $\varepsilon_l$
4. End for
5. 找出错误率最小的混合特征 $\delta$ 作为最终的混合特征

由于本文使用的特征混合策略只需要遍历就能实现, 弱分类器也只需要简单查询的方法就能建立, 因此, 组合特征提取的速度要高于使用 Adaboost 学习的提取方法。为了便于表示与运算, 本文对混合特征进行了编码, 假定一组二进制数  $(01111)_2$  分别对应  $\{\xi_i\}$  中按采集顺序排列的 5 个子块特征, 当为 0 时表示对应的子块没有被选中, 为 1 时表示子块被选中, 将  $\{\xi_i\}$ ,  $\{\psi_i\}$  和  $\{\zeta_i\}$  中的对应二进制值组合起来形成的编码就是混合特征编码, 如图 7:

$\{\xi_i\} \rightarrow (01111)_2 = (15)_{10}$

$\{\zeta_i\} \rightarrow (010)_2 = (2)_{10}$

$\{\psi_i\} \rightarrow (01110)_2 = (14)_{10}$

编码

$\rightarrow \delta_l$

$(15, 2, 14)_{10}$

图 7 混合特征编码

4 Adaboost 学习

我们对判定树弱分类器使用 Adaboost 算法来生成强分类器。目前 Adaboost 比较常用的算法有两种: Realboost 和 Gentleboost, 文献 [14] 中提出了一种 ModestAdaBoost 方法, 并证明了该方法是优于 Realboost 和 Gentleboost 的。因此, 本文中使用 ModestAdaBoost 方法来进行强分类器的生成。由于 AdaBoost 是典型的 2 类分类器, 因此, 我们在训练一个

表情种类的分类器时,将其他所有种类的表情都作为负样本。具体算法如下:

表 4 Modest AdaBoost 算法流程

算法 3 Modest AdaBoost 算法流程
1. 给定训练数据集 $(x_1, y_1), \dots, (x_N, y_N)$ , 其中 $y_i \in \pm 1$ , 并初始化数据权重: $D_0(i) = 1/N, i = 1, \dots, N$
2. 运行算法 2, 从特征组合中找出错误率最小的组合特征
3. for $t = 1, \dots, T$ , 并且 $f_m \neq 0$ ( $T$ 表示 boost 迭代次数)
4. 运行算法 1 训练一个基于组合特征的按照分布 $D_t(i)$ 采样的弱分类器 $h_t(x)$
5. 计算 $D_t(i)$ 的“逆”分布 $\bar{D}_t(i) = (1 - D_t(i))\bar{Z}_t$ $\bar{Z}_t$ 表示 $Z_t$ 的“逆”
6. 计算基于不同 $\pm 1$ 标签值的概率密度 $P_{D_t}, D_t$ 表示输入数据的权重分布: $P_t^+(x) = P_{D_t}(y = +1 \cap h_t(x))$ $\bar{P}_t^+(x) = P_{\bar{D}_t}(y = +1 \cap h_t(x))$ $P_t^-(x) = P_{D_t}(y = -1 \cap h_t(x))$ $\bar{P}_t^-(x) = P_{\bar{D}_t}(y = -1 \cap h_t(x))$ 令 $f_t(x) = (P_t^+(1 - \bar{P}_t^+) - P_t^-(1 - \bar{P}_t^-))(x)$ , 同时更新分布: $D_{t+1}(i) = \frac{D_t(i)\exp(-yf_t(x_i))}{Z_t}$ $Z_t$ 表示归一化系数
7. End for
8. 建立最终的分类器: $F(x) = \text{sign}[\sum_{t=1}^T f_t(h_t(x))]$

5 实验

本文实验所使用的数据库是 Cohn - Kanade 表情数据库<sup>[3]</sup>, 该数据是目前使用最多的评估表情识别算法的数据库, 库中包含 100 名 18 ~ 30 岁的学生, 其中女性学生占 65%, 每一个学生的表情都包含一组 23 帧的图像序列。由于该库中有些学生的表情表达并不明显, 且每一个图像序列中最后三帧图像的表情最为明显的, 因此, 本文在该数据库中选取了 94 名表情表达非常明显的学生, 共 339 组表情图像序列中的最后 3 帧图像作为表情样本, 涵盖的表情种类有: 愤怒、厌恶、恐惧、高兴、悲伤和惊讶, 如图 8 所示。

与文献 [10] 和 [11] 中的数据一样, 本文对样本



图 8 表情非常明显的图像

按照双眼间距离将所有图像归一化到  $64 \times 64$  个像素大小。提取面部运动单元子块的大小为  $16 \times 16$  像素, 每个子块中提取的 Haar 特征数为 15 219 个, 因此, 总共 13 个子块总共提取了 197 847 个 Haar 特征。

实验中, 本文随机抽取了表情样本中 62 个学生的表情作为训练样本, 将其余的 32 个学生的表情作为测试样本, 得出了下面的数据, 并与文献 [10] 中的结果作了比较。

表 5 对表情最为明显图像的识别结果

表情类别	文献 [10] 的结果	本文的识别结果	本文组合特征的编码
愤怒	84.38	93.16	(15, 2, 14)
恐惧	93.18	89.46	(15, 1, 14)
高兴	95.59	94.2	(27, 4, 14)
厌恶	91.67	94.02	(30, 2, 14)
悲伤	90.38	97.44	(30, 1, 14)
惊讶	98.33	98.8	(15, 1, 14)
平均正确率 (%)	92.3	94.52	

从上述结果看, 本文的方法在识别表情最为明显图像时的效果要优于文献 [10] 中的方法。本文在提取面部运动单元的特征时, 加入了人眼和口部的粗定位, 虽然定位的偏差可能反过来影响识别结果, 但是最终结果还是要比从大量局部特征中采用耗时的机器学习方法找成分特征的结果要好一些。

在此基础上, 本文接着从 94 名学生, 共 339 组表情图像序列中选取了 6 帧图像, 其中 3 帧为表情表达最为明显的图像, 其余 3 帧为表情表达不明显的图像, 如图 9 所示。



图 9 表情不是很明显的图像

在此训练集进行实验, 最终平均识别正确率仅为 85.2%, 识别率下降的非常明显, 如何提高算法的鲁棒性, 也是下一步需要研究的问题。

## 6 总结

本文通过采用一种更为简单的混合特征提取方法,并结合 Modest Adaboost 方法,提出了一种基于面部运动单元的表情识别方法,通过在 Kohn-Canada 数据库上的实验,取得了良好的识别率。但是,仍然存在以下不足: (1) 本文虽然在提取混合特征的步骤上非常快捷,但是最后使用 ModestAdaboost 构建强分类器时特征向量维数过大; (2) 本文将该方法用于识别表情表达并不明显的图像,最终的结果并不理想。以上两点都是本文需要继续改进的地方。

## 参考文献(References)

- [1] Kman P E, Friesen W V. Facial action coding system [M]. Palo Alto, California: Consulting Psychologists Press, 1978.
- [2] Shan Caifeng, Gong Shaogang, McOwan P W. Robust facial expression recognition using local binary patterns [C]//2005 IEEE International Conference on Image Processing. 2005: II 370 – 373.
- [3] Kanade T, Cohn J, Tian Y L. Comprehensive database for facial expression analysis [C]//Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition( FG'00) . 2004: 46 – 53.
- [4] Littlewort G, Bartlett M S, Fasel I, et al. Movellan J. Dynamics of facial expression extracted automatically from video [J]. Image and Vision Computing, 2006, ( 24 ) : 615 – 625.
- [5] 朱健翔, 苏光大, 李迎春. 结合 Gabor 特征与 Adaboost 的人脸表情识别 [J]. 光电子 · 激光, 2006, 17( 8 ) : 993 – 998.
- ZHU Jianxiang, SU Guangda, LI Yingchun. Facial Expression Recognition Based on Gabor Feature and Ada-boost [J]. Optoelectronics · Laser, 2006, 17( 8 ) : 993 – 998. ( in Chinese )
- [6] 王映辉. 人脸识别: 原理、方法与技术 [M]. 北京: 科学出版社, 2010: 245 – 258.
- [7] 王志良, 孟秀艳. 人脸工程学 [M]. 北京: 机械工业出版社, 2008: 49 – 52.
- [8] Ding Liya, Martinez A M. Features versus context: an approach for precise and detailed detection and delineation of faces and facial features [J]. IEEE Transactions On Pattern Analysis and Machine Intelligence, 2010, 38 ( 11 ) : 2022 – 2038.
- [9] Simon T, Nguyen M H, Torre F D L, et al. Action unit detection with segment-based SVMs [C]//IEEE International Conference on Computer Vision and Pattern Recognition. San Francisco, 2010: 2737 – 2744.
- [10] Yang Peng, Liu Qingshan, Metaxas D N. Exploring facial expressions with compositional features [C]//IEEE International Conference on Computer Vision and Pattern Recognition. 2010: 2638 – 2644.
- [11] Yang P, Liu Q, Metaxas D N. Facial expression recognition using encoded dynamic features [C]//2008 IEEE Conference on Computer Vision and Pattern Recognition. 2008: 1 – 8.
- [12] Viola P, Jones M J. Robust real-time face detection [J]. International Journal of Computer Vision, 2004, 57( 2 ) : 137 – 154.
- [13] Duda R O, Hart P E, Stork D G. 模式分类 [M]. 李宏东, 译. 北京: 机械工业出版社, 2003.
- Duda R O, Hart P E, Stork D G. Pattern classification [M]. LI Hongdong, trans. Beijing: China Machine Press, 2003. ( in Chinese )
- [14] Vezhnevets A, Vezhnevets V. Modest AdaBoost: teaching AdaBoost to generalize better [C]//Graphicon 2005 Proceedings. Russia: 2005.