

基于支持向量机的人脸检测训练集增强*

王瑞平^{1,2,3}, 陈杰⁴, 山世光^{1,2+}, 陈熙霖^{1,2}, 高文^{2,5}

¹(中国科学院 智能信息处理重点实验室, 北京 100190)

²(中国科学院 计算技术研究所, 北京 100190)

³(中国科学院 研究生院, 北京 100049)

⁴(哈尔滨工业大学 计算机科学与技术学院, 黑龙江 哈尔滨 150001)

⁵(北京大学 信息科学技术学院, 北京 100871)

Enhancing Training Set for Face Detection Based on SVM

WANG Rui-Ping^{1,2,3}, CHEN Jie⁴, SHAN Shi-Guang^{1,2+}, CHEN Xi-Lin^{1,2}, GAO Wen^{2,5}

¹(Key Laboratory of Intelligent Information Processing, The Chinese Academy of Sciences, Beijing 100190, China)

²(Institute of Computing Technology, The Chinese Academy of Sciences, Beijing 100190, China)

³(Graduate University, The Chinese Academy of Sciences, Beijing 100049, China)

⁴(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

⁵(School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China)

+ Corresponding author: E-mail: sgshan@jdl.ac.cn, <http://www.jdl.ac.cn>

Wang RP, Chen J, Shan SG, Chen XL, Gao W. Enhancing training set for face detection based on SVM. *Journal of Software*, 2008,19(11):2921–2931. <http://www.jos.org.cn/1000-9825/19/2921.htm>

Abstract: According to support vector machines (SVMs), for those geometric approach based classification methods, examples close to the class boundary usually are more informative than others. Taking face detection as an example, this paper addresses the problem of enhancing given training set and presents a nonlinear method to tackle the problem effectively. Based on SVM and improved reduced set algorithm (IRS), the method generates new examples lying close to the face/non-face class boundary to enlarge the original dataset and hence improve its sample distribution. The new IRS algorithm has greatly improved the approximation performance of the original reduced set (RS) method by embedding a new distance metric called image Euclidean distance (IMED) into the kernel function. To verify the generalization capability of the proposed method, the enhanced dataset is used to train an AdaBoost-based face detector and test it on the MIT+CMU frontal face test set. The experimental results show that the original collected database can be enhanced effectively by the proposed method to learn a face detector with improved generalization performance.

Key words: face detection; SVM (support vector machine); reduced set method; AdaBoost

* Supported by the National Natural Science Foundation of China under Grant Nos.60533030, 60673091, 60772071 (国家自然科学基金); the National High-Tech Research and Development Plan of China under Grant Nos.2006AA01Z122, 2007AA01Z163 (国家高技术研究发展计划(863)); the 100 Talents Program of the Chinese Academy of Sciences (中国科学院“百人计划”); the ISVISION Technology Co. Ltd. (上海银晨智能识别科技有限公司)

Received 2007-04-17; Accepted 2007-08-03

摘要: 根据支持向量机(support vector machine,简称 SVM)理论,对基于边界的分类算法(geometric approach)而言,类别边界附近的样本通常比其他样本包含有更多的分类信息.基于这一基本思路,以人脸检测问题为例,探讨了对给定训练样本集进行边界增强的问题,并为此而提出了一种基于支持向量机和改进的非线性精简集算法 IRS(improved reduced set)的训练集边界样本增强算法,用以扩大训练集并改善其样本分布.其中,所谓 IRS 算法是指在精简集(reduced set)算法的核函数中嵌入一种新的距离度量——图像欧式距离——来改善其迭代近似性能,IRS 可以有效地生成新的、位于类别边界附近的虚拟样本以增强给定训练集.为了验证算法的有效性,采用增强的样本集训练基于 AdaBoost 的人脸检测器,并在 MIT+CMU 正面人脸测试库上进行了测试.实验结果表明,通过这种方法能够有效地提高最终分类器的人脸检测性能.

关键词: 人脸检测;支持向量机;精简集算法;AdaBoost

中图法分类号: TP391 **文献标识码:** A

在过去的 10 多年中,人脸检测在计算机视觉领域得到了广泛的应用,并引起了越来越多的学者的深入研究.人脸检测是指给定一幅图像,判定其中是否存在人脸,如果有人脸,则返回其坐标和大小^[1].近年来,人脸检测领域研究的重点集中于数据驱动的基于统计学习的方法^[1,2].Osuna 等人首次提出了一种采用支持向量机学习人脸检测器的方法^[3].Sung 和 Poggio 等人提出了一种基于“高斯簇”对人脸和非人脸的分布进行建模的算法^[4].Rowley 等人提出了一种基于神经网络的人脸检测系统^[5].Schneiderman 和 Kanade 提出了一种基于后验概率函数估计的人脸检测器^[6].Yang 等人提出了一种基于稀疏神经网络(sparse network of winnow,简称 SNoW)的学习体系结构^[7].最近,一种流行的算法是 Viola 提出的基于 Boosting 的快速人脸检测系统^[8].在此之后,一些学者对基于 Boosting 的算法提出了改进^[9-14].此外,Jin 等人将改进的 LBP(local binary pattern)算子与贝叶斯框架融合,学习人脸检测器,对于光照变化较大的图像有较好的检测效果^[15].Fang 等人提出了一种融合人脸轮廓信息和面部区域信息的人脸检测器,有效地提高了人脸检测算法的精度^[16].

对基于统计模型的人脸检测系统来说,训练或测试一个分类器都要进行数据的收集,最终学习到的分类器性能在很大程度上依赖于所收集的训练集合样本的分布.支持向量机的理论表明,分布在类别边界上的样本(即支持向量)将在基于边界的分类器学习算法中发挥重要作用^[17].目前,另一种得到广泛应用的学习算法是 AdaBoost.该算法的一个主要特点是,通过对错分的样本加大权重来实现样本重采样,观察发现那些错分的样本大多数分布在决策边界附近.由此可以看出,对基于边界的分类算法(geometric approach)而言,分布在类别边界附近的样本通常比那些距离边界较远的样本包含更多的分类判别信息,从而对分类器的学习影响也更大.因此,如何充分而有效地利用所收集的训练集来学习具有更好泛化能力的分类器显得非常重要,而这方面相关的研究还比较少.Lu 和 Jain 提出一种对已有样本进行重采样的算法^[18].Zhou 和 Jiang 利用神经网络集成(neural network ensemble)产生虚拟的训练样本集,以学习具有更强推广性能的决策树分类器^[19].他们对分类器的错误率构成、训练集数据分布和训练数据中噪声的影响等进行了分析,从而在理论上证明了生成虚拟样本可以有效地辅助机器学习.此外,他们还给出了实验的验证结果.Chen 等人提出了一种基于遗传算法来进行人脸检测训练集样本扩张的算法^[20],通过生成新的虚拟人脸本来模拟人脸的一些变化,比如不同的光照、图片质量等,从而改善训练集的分布.该算法对基于分布的分类器(density based approach)来说可能会取得较好的结果,但是该算法没有专门研究那些处于人脸与非人脸类别边界的样本.本文的目的是产生类别边界上的虚拟样本,同时包括人脸和非人脸样本.

基于上述考虑,我们提出了基于精简集方法(reduced set method)和图像欧式距离 IMED(image Euclidean distance)的非线性算法^[21],该算法通过产生边界上的虚拟本来增强收集的训练集.本文在此基础上作了进一步的讨论:针对嵌入 IMED 的改进型精简集方法,本文进一步探讨了其他图像距离度量的可用性,并通过相关的实验进行了性能对比,验证了采用 IMED 距离的合理性;针对精简集算法生成的虚拟样本数量选取的问题,本文通过实验对比给出了一个较为合理的样本数量选取准则.

本文第 1 节阐述如何采用基于 IMED 的改进型精简集算法生成虚拟边界样本以增强样本集.第 2 节描述该

算法的实验结果.第 3 节给出相关的结论.

1 基于IMED的改进型精简集算法

本节首先简要介绍精简集算法和图像欧式距离 IMED,然后提出基于 IMED 的改进型精简集算法 IRS (improved reduced set).系统的流程如图 1 所示,先将原始集合分为正例(人脸)和反例(非人脸)样本集来学习 SVM 分类器;采用 IRS 算法,针对分类器学习得到的两类支持向量分别生成各自类别的精简集向量(即虚拟的边界样本);最后,将生成的两类虚拟样本加入各自类别的原始训练集合,即可得到增强的训练样本集,用于学习最终的分

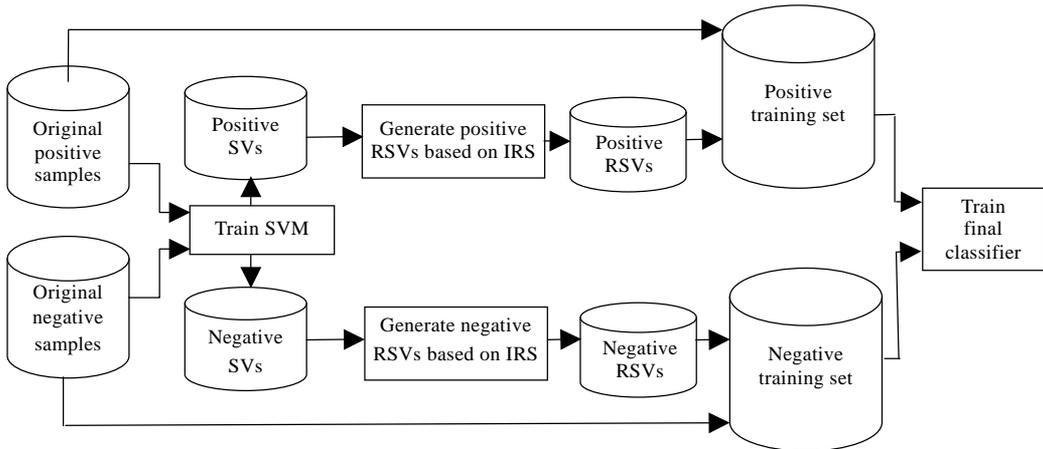


Fig.1 Flow chart of the proposed method

图 1 本文算法流程图

1.1 精简集算法

基于 SVM 的分类器在人脸检测和相关的模式识别领域中已经得到成功的应用^[3,22],但是它们存在一个共同的问题就是在测试阶段速度较慢.为此,Burges 和 Schölkopf 等人提出了精简集算法^[22-24],从而将 SVM 的分类器在测试阶段的速度提高了数十倍以上,同时保证分类器的精度仅有微小损失.首先简单回顾一下 SVM 分类器的基本原理.一般来说,SVM 隐式地将原始低维空间中的样本 $(x_1, y_1), \dots, (x_l, y_l) \in X \times \{\pm 1\}$ 映射到一个高维(其维数可能为无穷大)内积特征空间中,这一映射过程通常通过一个非线性函数 $\Phi: X \rightarrow F, x \mapsto \Phi(x)$ 来实现,这里 x_i 表示原始空间中的训练样本, y_i 为相应的类别标号.在特征空间 F 中,SVM 计算具有最大边界间隔(margin)的分类超平面,使得距离该超平面最近的两类待分样本的间隔最大.通常可以采用核函数 $k(x, x')$ 来计算特征空间中的内积,即 $k(x, x') = (\Phi(x) \cdot \Phi(x'))$.SVM 分类器学习得到的最终判别函数形式为

$$f(x) = \text{sgn}\left(\sum_{i=1}^l \alpha_i y_i k(x, x_i) + b\right) \tag{1}$$

其中, $0 \leq \alpha_i \leq C, i = 1, \dots, l, \sum_{i=1}^l \alpha_i y_i = 0$ (常数 C 是训练过程中边界间隔最大化与训练误差最小化之间的平衡因子).对应于系数 $\alpha_i > 0$ 的训练样本 x_i 即被称为支持向量(support vector, 简称为 SV).显然,在测试阶段,分类器的速度与支持向量的个数成正比.常用的核函数包括多项式核函数、高斯(Gaussian)核函数以及 Sigmoid 核函数等.由于高斯核函数具有性能稳定、参数少等优点^[25],本文采用该核函数计算特征空间的内积,其形式如下:

$$k(x, y) = \exp(-\|x - y\|^2 / 2\sigma^2) \tag{2}$$

给定特征空间的一个向量 $\Psi \in F$,其表示形式为原始空间的一组向量在特征空间的投影向量之和,

$$\Psi = \sum_{i=1}^{N_x} \alpha_i \Phi(x_i) \tag{3}$$

这里, $\alpha_i \in \mathbb{R}, \mathbf{x}_i \in \mathbf{X}$. 为了降低计算复杂度, 精简集算法寻找一组新的向量展开式来近似 Ψ , 并将这组新的向量展开式记作

$$\Psi' = \sum_{i=1}^{N_z} \beta_i \Phi(\mathbf{z}_i) \quad (4)$$

其中, $N_z \ll N_x, \beta_i \in \mathbb{R}, \mathbf{z}_i \in \mathbf{X}$ 称为精简集向量(reduced set vector, 简称 RSV). 这样, 最小化近似误差的准则可以表示为如下目标函数^[22]:

$$\varepsilon = \|\Psi - \Psi'\|^2 = \sum_{i,j=1}^{N_x} \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) + \sum_{i,j=1}^{N_z} \beta_i \beta_j k(\mathbf{z}_i, \mathbf{z}_j) - 2 \sum_{i=1}^{N_x} \sum_{j=1}^{N_z} \alpha_i \beta_j k(\mathbf{x}_i, \mathbf{z}_j) \quad (5)$$

该算法的关键思想在于, 虽然映射函数 Φ 的形式没有显式给出, 但是上述目标函数式仍可以通过核函数计算得到. 文献[23]提出了一种有效的迭代算法来计算公式(5)中的向量 \mathbf{z}_i 与系数 β_i .

对于给定的精简集向量个数 N_z , 将公式(4)代入公式(1)中即可得到:

$$f'(\mathbf{x}) = \text{sgn}\left(\sum_{j=1}^{N_z} \beta_j k(\mathbf{x}, \mathbf{z}_j) + b\right) \quad (6)$$

采用公式(6)中计算得到的分类器 $f'(\mathbf{x})$ 作为对原始训练的 SVM 判别函数式(1)的近似, 可以保证分类器精度的损失尽量小. 阈值参数 b 可以通过在单独的校验集上调节以获得最佳性能^[24]. 但是, 由于 $N_z \ll N_x$, 它能够极大地提高测试阶段的计算效率. 文献[24]将该算法应用于人脸检测, 取得了很好的效果.

1.2 图像欧式距离

与传统的欧式距离相比, 图像欧式距离考虑到了图像中像素点之间的空间位置关系, 它具有以下特点: 对轻微的图像扰动不敏感, 计算简单, 易于嵌入到大多数图像识别技术中^[26].

给定两幅大小为 $M \times N$ 的图像 \mathbf{x}, \mathbf{y} , 它们之间的图像欧式距离 $d_{IMED}(\mathbf{x}, \mathbf{y})$ 由以下公式给出:

$$d_{IMED}^2(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^T \mathbf{G} (\mathbf{x} - \mathbf{y}) \quad (7)$$

这里, $\mathbf{G} = (g_{ij})_{MN \times MN}$ 是一个 MN 阶的对称半正定矩阵(当 \mathbf{G} 取单位矩阵时, $d_{IMED}(\mathbf{x}, \mathbf{y})$ 退化为传统的欧式距离). 通常可以采用高斯函数来定义矩阵 \mathbf{G} 的元素 g_{ij} :

$$g_{ij} = f(|P_i - P_j|) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{|P_i - P_j|^2}{2\sigma^2}\right) \quad (8)$$

其中, σ 是高斯窗宽参数(本文实验中取值为 1); $P_i, P_j (i, j = 1, 2, \dots, MN)$ 代表像素点, $|P_i - P_j|$ 为图像网格中这两个像素点位置之间的距离.

为了提高实际使用中该算法的计算效率, 文献[26]中采用了一种快速计算方法. 因为矩阵 \mathbf{G} 的计算与具体的图像数据内容无关, 只与图像大小有关, 所以可以事先计算好. 具体的方法是将矩阵 \mathbf{G} 分解为 $\mathbf{G} = \mathbf{G}^{1/2} \mathbf{G}^{1/2}$, $\mathbf{G}^{1/2} = \mathbf{\Gamma} \mathbf{\Lambda}^{1/2} \mathbf{\Gamma}^T$, 其中, $\mathbf{\Lambda}$ 与 $\mathbf{\Gamma}$ 分别是矩阵 \mathbf{G} 的特征值和特征向量矩阵. 实际使用时, 可以先将原始图像 \mathbf{x}, \mathbf{y} 变换为 $\mathbf{u} = \mathbf{G}^{1/2} \mathbf{x}, \mathbf{v} = \mathbf{G}^{1/2} \mathbf{y}$, 这样, \mathbf{x}, \mathbf{y} 的 IMED 距离即可转化为变换图像 \mathbf{u}, \mathbf{v} 之间的传统欧式距离, 即

$$(\mathbf{x} - \mathbf{y})^T \mathbf{G} (\mathbf{x} - \mathbf{y}) = (\mathbf{x} - \mathbf{y})^T \mathbf{G}^{1/2} \mathbf{G}^{1/2} (\mathbf{x} - \mathbf{y}) = (\mathbf{u} - \mathbf{v})^T (\mathbf{u} - \mathbf{v}) \quad (9)$$

上述变换 $\mathbf{G}^{1/2}(\cdot)$ 称为标准化变换(standardizing transform, 简称 ST). 可以看出, 将变换后的图像作为其他算法的输入即相当于在这些算法中自动嵌入了 IMED 距离度量.

1.3 改进的精简集算法(IRS)

如第 1.1 节所述, 精简集算法可以提高 SVM 分类器的测试速度. SVM 判别函数中的权向量 $\mathbf{w} = \sum_{i=1}^{N_x} \alpha_i \mathbf{y}_i \Phi(\mathbf{x}_i)$ ($\alpha_i > 0, \mathbf{y}_i = \pm 1; N_x$ 为支持向量的个数) 可以通过计算展开式 $\mathbf{w}' = \sum_{j=1}^{N_z} \beta_j \Phi(\mathbf{z}_j)$ (其中, \mathbf{z}_j 是一组向量, β_j 是相应的系数) 来近似. 由于支持向量 \mathbf{x}_i 与精简集向量 \mathbf{z}_j 属于同一个空间 \mathbf{X} , 因此, 模式 \mathbf{z}_j 的类别可由其对应的系数 β_j 决定, 即对应系数为正的精简集向量属于正例样本, 反之属于反例样本^[23].

本文提出的改进型精简集算法 IRS 对权向量 \mathbf{w} 进行分解, 即将正例与反例(本文为人脸与非人脸类别)支持向量分别组织在一起, 如下式:

$$\mathbf{w} = \mathbf{w}^p + \mathbf{w}^n \quad (10)$$

其中,

$$\mathbf{w}^p = \sum_{k=1}^{N_x^p} \alpha_k^p \Phi(\mathbf{x}_k^p), \mathbf{w}^n = \sum_{l=1}^{N_x^n} \alpha_l^n \Phi(\mathbf{x}_l^n), N_x = N_x^p + N_x^n \quad (11)$$

系数 $\alpha_k^p > 0$ 对应于正例的支持向量 \mathbf{x}_k^p , 系数 $\alpha_l^n < 0$ 对应于反例的支持向量 \mathbf{x}_l^n . N_x^p 与 N_x^n 分别为两类支持向量的个数.

原始的精简集算法是把向量 \mathbf{w} 看作一个整体来求其近似向量, 这一近似过程在某种意义上相当于在两类样本空间中同时进行向量聚类运算^[23]. 本文的目标在于产生虚拟的两类边界样本来增强初始样本集, 从而提高最终分类器的判别性能. 为了方便控制产生的两类虚拟样本的数量并提高向量近似运算的迭代收敛速度, IRS 算法对两个向量 \mathbf{w}^p 与 \mathbf{w}^n 分别进行近似计算, 即相当于分别在两类样本内部单独进行文献[23]所述的向量聚类运算. \mathbf{w}^p 与 \mathbf{w}^n 的近似计算如下:

$$\mathbf{w}^{p'} = \sum_{m=1}^{N_x^p} \beta_m \Phi(\mathbf{z}_m), \mathbf{w}^{n'} = \sum_{n=1}^{N_x^n} \beta_n \Phi(\mathbf{z}_n) \quad (12)$$

由于向量 $\mathbf{z}_m(\mathbf{z}_n)$ 的类别标签由其相应的系数 $\beta_m(\beta_n)$ 来决定, 本文实验表明, 向量 $\mathbf{w}^{p'}$ 展开式中的所有系数 β_m 均为正值, 而 $\mathbf{w}^{n'}$ 展开式中的所有系数 β_n 均为负值. 这意味着对两类样本分别进行向量聚类运算时, 得到的一系列精简集向量均属于相应的类别标签, 这一结果也符合直观上的理解.

精简集算法的一个重要指标是其近似性能. 文献[23]指出, 在求向量近似运算的过程中, 可以通过采用不同初始值来降低数值计算不稳定问题对迭代算法收敛的影响. 然而, 仅仅采用改变迭代初始值这种策略对于提高算法的近似性能效果不明显. 在 IRS 算法中, 通过在迭代生成精简集向量的高斯核函数(公式(2))中嵌入图像欧式距离 IMED, 可以有效地改进算法的近似性能. 为此, 只需将 $\|\mathbf{x}-\mathbf{y}\|^2$ 替换为 $(\mathbf{x}-\mathbf{y})^T \mathbf{G}(\mathbf{x}-\mathbf{y})$, 并代入公式(9)即可得到最终的核函数:

$$k_{IMED}(\mathbf{x}, \mathbf{y}) = \exp\left(\frac{-(\mathbf{x}-\mathbf{y})^T \mathbf{G}(\mathbf{x}-\mathbf{y})}{2\sigma^2}\right) = \exp(-\|\mathbf{u}-\mathbf{v}\|^2 / 2\sigma^2) \quad (13)$$

除了图像欧式距离 IMED 之外, 其他一些代表性的图像距离度量有切距离^[27]、广义 Hausdorff 距离^[28]等. 其中, 切距离对于一些特定的图像变换具有局部不变性, 并且在手写体数字识别中得到了广泛的应用^[27,29]; 广义 Hausdorff 距离不仅对噪声鲁棒, 而且可以用来比较两幅图像局部区域的相似性, 已经成为形状比较的标准度量工具. 然而, 这两种距离的计算复杂度高, 而且难以嵌入到其他识别分类算法中, 从而不适合用来实现改进精简集算法近似性能的目标. 同时, 文献[26]的实验表明, 对于无噪声图像, 采用 IMED 距离取代传统的欧式距离仍然可以稳定地提高识别算法的性能. 因此, 本文将 IMED 距离嵌入到精简集算法中, 以提高其近似性能. 下一节的实验中将以切距离为例, 考察其与 IMED 距离的性能效率对比以及对算法的适用性.

2 实验结果

本节主要有两组实验: 第 1 组实验用于验证 IMED 距离的有效性, 实验数据来自于 MIT 数据库; 第 2 组实验用于验证算法所生成的虚拟样本对于提升分类器性能的有效性, 实验数据来自于手工收集的一个数据集.

2.1 不同图像距离度量嵌入精简集算法的比较

2.1.1 3 种距离度量

常用的图像距离度量包括欧式距离、切距离、Hausdorff 距离和图像欧式距离 IMED. 由于 Hausdorff 距离主要用于计算两个二值图像集合之间的距离, 不适用于本文讨论的算法, 故本节的实验主要对比另外 3 种距离度量方式.

分别将 3 种距离度量嵌入精简集算法的核函数中. 采用传统欧式距离的算法(即为原始的精简集算法, 以下称为 ED-RSV), 核函数即为公式(2); 采用 IMED 距离的算法(以下称为 IMED-RSV), 核函数为 $k_{IMED}(\mathbf{x}, \mathbf{y})$; 而采用切距离的算法(以下称为 TD-RSV), 核函数形式与公式(2)相同, 但其中距离度量 $\|\mathbf{x}-\mathbf{y}\|^2$ 改为两幅图像的切距离模式. 为了比较 3 种不同的距离在提高精简集算法的近似性能与计算效率方面的差别, 类似于文献[23,24], 将权重向量 \mathbf{w} 作为一个整体来近似.

2.1.2 实验结果

实验数据来源于 MIT 数据库^[30].该数据库包含训练集和测试集两个集合.训练集含有 6 977 幅图像(2 429 个人脸样本,4 548 个非人脸样本),测试集含有 24 045 幅图像(472 个人脸样本,23 573 个非人脸样本).所有样本均归一化为 20×20 大小的灰度图像.

首先采用训练集提供的 6 977 个样本学习基于 SVM 的分类器.在学习之前,这些样本先进行直方图均衡化预处理,然后利用图像的 400 维灰度值向量作为样本特征.核函数采用式(2)所示的高斯核函数,窗宽参数 σ 值通过交叉验证设为 5(后续精简集算法中的高斯核函数也采用相同的 σ 值).训练得到的分类器的支持向量个数为 1 148,其中包括 475 个人脸样本和 673 个非人脸样本.

在产生精简集向量序列的算法中,一个重要的问题是终止条件的设置.本文中采用如下准则:

$$J=|\varepsilon_{i+1}-\varepsilon_i|/\varepsilon_i \quad (14)$$

这里, $|\varepsilon_{i+1}-\varepsilon_i|$ 表示连续产生的两个向量对应近似误差的差别,参数 J 值随产生的精简集向量个数增多而下降.实验表明,当参数 $J<0.001$ 时,近似误差下降曲线趋于平缓,如图 2 所示.此时,新生成的向量仅能带来有限的近似性能改进,因此,实验中将终止条件设为 $J<0.001$.此外,为了进行对比,实验中需要保证 3 种算法产生同等数量的精简集向量,在该数量下 3 种算法均达到终止条件要求.最终得到 3 组精简集向量,每组含 300 个样本.针对 3 种算法 ED-RSV,IMED-RSV 和 TD-RSV,相应的正例样本(人脸图像)数量依次为 100,102,105;相应的反例样本(非人脸图像)数量依次为 200,198,195.

3 种算法相应的近似误差下降曲线对比如图 2 所示.可以发现,通过在精简集算法中嵌入 IMED 距离与切距离来取代传统的欧式距离,都可以稳定地提高算法的近似性能.主要原因在于,与传统欧式距离相比,IMED 距离和切距离均考虑了图像中像素之间的相关性,并且对图像中较为轻微的变换扰动具有一定的鲁棒性,从而能够在一定程度上消除噪声带来的影响.同时,算法 TD-RSV 的近似性能略优于 IMED-RSV,可能的原因在于切距离的计算比 IMED 距离更为复杂,且对图像噪声更加鲁棒.

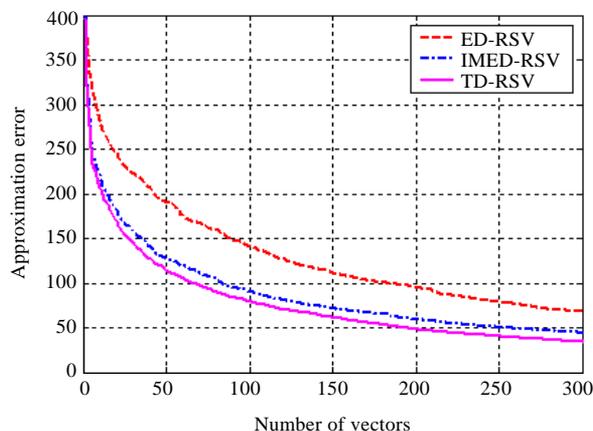


Fig.2 Approximation error as a function of the number of generated reduced set vectors

图 2 近似误差随产生的精简集向量数量变化的曲线

图 3 给出了 3 种算法产生精简集向量所进行的迭代次数统计.从中可以看出,算法 IMED-RSV 与 TD-RSV 所需的平均迭代次数均高于原始的 ED-RSV 算法,其可能原因在于,这两种嵌入了新的复杂距离度量的算法对迭代过程中生成的中间向量也都要考虑与之相关的图像扰动和变换等因素,从而使收敛过程变慢.

对比 3 种算法的运行效率,ED-RSV 算法的计算最为快速、简单,效率也最高;TD-RSV 算法中切距离的计算较为复杂、耗时.在 IMED-RSV 算法中,IMED 距离的计算采用文献[26]的快速算法,只需将两幅图像进行相应的标准化变换(ST),再计算变换后图像的传统欧式距离即可,而且图像只需做一次标准化变换就可以保证算法多次迭代的需要,这使得 IMED-RSV 算法可以达到接近于 ED-RSV 算法的计算效率.而切距离的计算则不具备

这个优势,每次迭代都需要计算不同图像样本向量之间的切距离,其计算复杂性很高.综合上述性能与效率两方面的分析比较,本文提出的改进型精简集算法最终采用嵌入 IMED 距离的核函数来提高算法的近似性能.

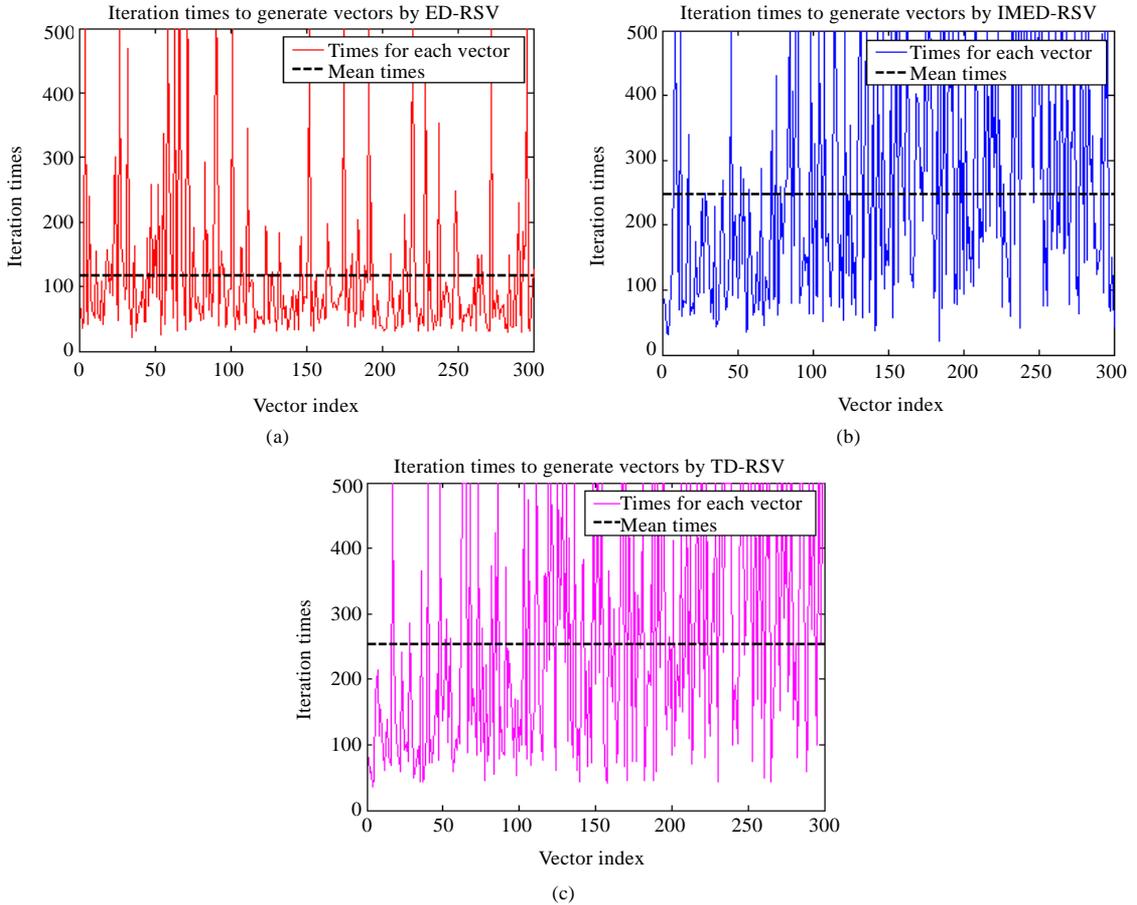


Fig.3 Iteration times to generate reduced set vectors by the three methods

图 3 3 种算法产生精简集向量所进行的迭代次数

图 4 形象地给出实验结果示意图(图中圆形代表人脸类别样本,方形代表非脸类别样本),图示中的边界样本包括 SVM 分类器学习到的支持向量与 IMED-RSV 算法生成的精简集向量,即虚拟的边界样本.可以看出,算法产生的虚拟样本与所属类别的真实样本几乎没有区别.

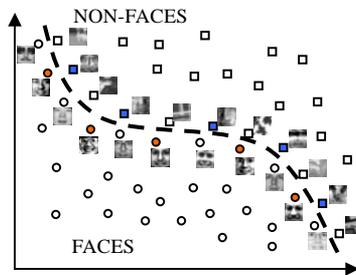


Fig.4 Some SVs (hollow) and RSVs (solid) lying on the boundary.

The RSVs are generated by the proposed IMED-RSV

图 4 类别边界上的支持向量(空心)与 IMED-RSV 算法生成的精简集向量(实心)

2.2 生成样本的有效性验证

2.2.1 分类器的训练

本节在一个较大规模的数据集上进行对比实验,目的是:1) 验证改进型精简集算法 IRS 所生成虚拟样本对于提升分类器性能的有效性;2) 考察虚拟样本的数量选取(即终止条件参数 J 的取值)对分类器性能的影响.该数据集包含 15 000 幅正面人脸图像和 25 000 幅非人脸图像,所有样本均归一化为 20×20 大小的灰度图像.由于在这个大数据集上运行精简集算法复杂度较高,实验中将其划分为 5 个同样大小的子集.

针对每个子集训练 SVM 分类器,将判别函数中的权向量 w 分解为正例支持向量集合 w^p 和反例支持向量集合 w^n ,采用 IRS 算法对 w^p 和 w^n 分别进行近似计算.最后,将各子集所得的正例精简集向量合并得到虚拟人脸样本集合,同时将各子集的反例精简集向量合并得到虚拟非人脸样本集合.表 1 给出了实验中依次设定的产生精简集向量序列的 3 个具体的终止条件与对应得到的两类虚拟样本数量.

Table 1 Stopping criteria and corresponding numbers of generated RSVs

表 1 不同的迭代终止条件与对应生成的精简集向量数量

Stopping criterion $\sup(\varepsilon_{t+1}-\varepsilon_t /\varepsilon_t)$	Faces	Nonfaces
0.01	350	617
0.001	910	1 535
0.000 5	1 515	2 520

分别将表 1 的 3 组虚拟样本加入到原始集合中得到 3 个增强的训练样本集,分别称为 Enhanced Set-1, Enhanced Set-2, Enhanced Set-3(注意到较小的集合依次是较大集合的子集).采用原始样本集与 3 个增强样本集分别训练基于 AdaBoost 的人脸检测器^[8].为使训练的检测器具有更好的实用性能,实验中采用 bootstrap 策略^[4]来增加学习分级分类器的非人脸样本,这些非人脸样本的提取来自 10 964 幅不含人脸的图像库.由于对训练的 4 个分类器均采用同样的策略,因此不会影响它们最终性能的对比结果.

2.2.2 测试结果

基于上一节的原始集合和 3 个增强的训练样本集训练得到 4 个基于 AdaBoost 的人脸检测器,然后将这 4 个检测器在 MIT+CMU 的正面人脸测试集上进行测试.该测试集共包括 130 幅图像,507 张正面向上的人脸^[5].测试结果如图 5 所示,从图中可以看出,采用本文提出的算法增强的样本集能够学习到性能明显更优的人脸检测器.具体地说,一方面,新生成的虚拟样本大大丰富了原始数据集,使得分类器能够学习到更多的类别边界信息,从而最终提升其推广性能.正如文献[19]所述,其理论分析也表明了利用产生的虚拟样本来提高分类器推广性能的合理性和有效性;另一方面,随着生成的新样本的不断加入,它们所能带来的有效边界信息逐渐减少,从而对提升分类器推广性能的作用逐渐减弱.从图 5 的 ROC 曲线可以看出,利用集合 Enhanced Set-3(对应于终止条件 $J < 0.0005$)训练的检测器相对于 Enhanced Set-2(对应于终止条件 $J < 0.001$)训练的检测器而言性能提高不明显.如第 2.1.2 节所述,当近似误差下降曲线开始趋于平缓时,虚拟样本的生成过程即可停止,因此,本文利用实验结果给出终止条件的一个经验值,即 $J < 0.001$.从图 5 中可以看到,采用增强的样本集合 3(Enhanced Set-3)训练的 AdaBoost 检测器的检测结果是:正确率为 90.73%,错误检测到的非人脸个数为 48.

为了进一步验证本文提出的样本集增强算法的有效性,接下来的实验采用上述原始样本集合与增强的样本集合 3(Enhanced Set-3)分别学习基于 SNoW^[7]与 SVM^[3]的人脸检测器,并在 MIT+CMU 正面人脸测试集上进行测试,测试参数采用与上述 AdaBoost 分类器完全相同的设置,即采用了相同的图像缩放因子、扫描窗口滑动步长和扫描窗口数量.ROC 曲线对比如图 6 所示.可以看出,针对基于边界的分类器学习算法(geometric approach,如 AdaBoost,SVM),采用本文算法增强的样本集训练所得的检测器具有明显的性能提升;而针对基于分布的分类器学习算法(density based approach,如 SNoW),增强的样本集所得检测器的性能提升则不明显.这是由算法本身的特点所决定的,如前所述,本文提出的改进型精简集算法 IRS 本质上是一种样本集边界增强算法,所生成的虚拟样本侧重于改进样本集类别边界上的分布.因此,本文提出的样本集增强算法对于基于边界的学习算法具有更好的适用性.

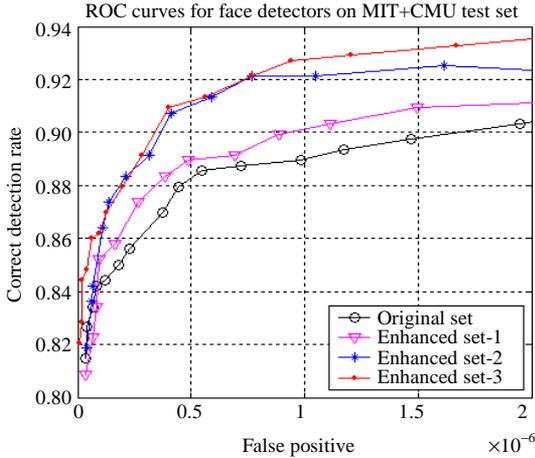


Fig.5 The ROC curves on the MIT+CMU test set by AdaBoost face detectors

图 5 AdaBoost 人脸检测器在 MIT+CMU 测试集上的 ROC 曲线对比

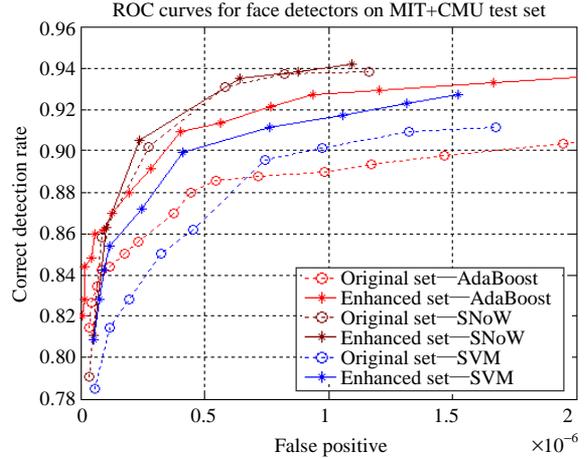


Fig.6 The ROC curves on the MIT+CMU test set by different detectors

图 6 不同的分类器学习算法在 MIT+CMU 测试集上的 ROC 曲线对比

表 2 列出了基于本文提出的算法和其他已发表的一些算法在 MIT+CMU 的正面人脸库中进行检测的结果比较.可以看出,本文算法训练的分类器基本上达到了与同类算法相当的检测效果;然而由于不同算法采用的训练集规模、测试参数都不尽相同,这里的性能比较只能提供一个相对的参考.图 7 给出一些在测试集上的检测实例.

Table 2 Detection rates comparison of our system and others

表 2 我们的系统和其他算法的检测率的比较

Methods	Schneiderman ^[6]	Viola ^[8]	Hou ^[13]	Brubaker ^[14]	Chen ^[20]	Our method
Detection rate (%)	94.4	88.5	93.9	94.3	93.69	90.73
False alarms	65	31	65	65	64	48

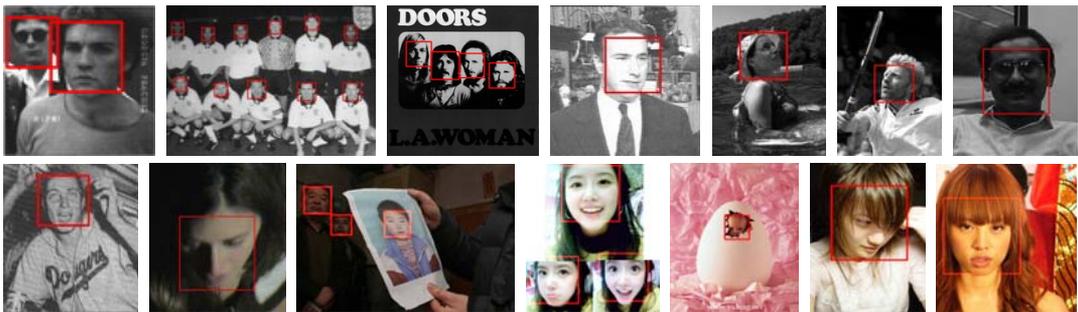


Fig.7 Some detection results of our trained detector

图 7 本文训练检测器的一些检测结果

3 结论与未来的工作

本文提出了一种基于支持向量机和改进的非线性精简集算法的样本集增强方法.该方法可以有效地生成位于类别边界处的样本,从而丰富原始的样本集合,增强给定训练集,改善训练样本分布.实验表明,与原始样本集相比,使用增强后的数据集训练的人脸检测器具有更好的检测性能.这一方面进一步印证了训练样本分布的

合理性对于统计学习方法最终性能的重要性,另一方面也表明本文所提出的方法确实可以有效地改善训练样本分布.本文可能的后续工作包括:(1) 在精简集算法中寻找其他更高效的距离度量;(2) 将所提出的算法扩展到多类分类问题中.

References:

- [1] Yang MH, Kriegman DJ, Ahuja N. Detecting faces in images: A survey. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2002,24(1):34–58.
- [2] Liang LH, Ai HZ, Xu GY, Zhang B. A survey of human face detection. *Chinese Journal of Computers*, 2002,25(5):449–458 (in Chinese with English abstract).
- [3] Osuna E, Freund R, Girosi F. Training support vector machines: An application to face detection. In: *Proc. of the CVPR*. 1997.
- [4] Sung KK, Poggio T. Example-Based learning for view-based human face detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1998,20(1):39–51.
- [5] Rowley HA, Baluja S, Kanade T. Neural network-based face detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1998,20(1):23–38.
- [6] Schneiderman H, Kanade T. A statistical method for 3D object detection applied to faces. In: *Proc. of the Computer Vision and Pattern Recognition*. Cambridge: IEEE Computer Society, 2000. 746–751.
- [7] Yang MH, Roth D, Ahuja N. A SNoW-based face detector. In: Solla SA, Leen TK, Müller KR, eds. *Proc. of the Advances in Neural Information Processing Systems 12*. Cambridge: MIT Press, 2000. 855–861.
- [8] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. In: *Proc. of the Conf. Computer Vision and Pattern Recognition (CVPR 2001)*. Cambridge: IEEE Computer Society, 2001. 511–518.
- [9] Xiao R, Zhu L, Zhang HJ. Boosting chain learning for object detection. In: *Proc. of the ICCV*. Washington: IEEE Computer Society, 2003. 709–715.
- [10] Li SZ, Zhu L, Zhang ZQ, Blake A, Zhang HJ, Shum H. Statistical learning of multi-view face detection. In: Heyden A, Sparr G, Nielsen M, Johansen P, eds. *Proc. of the 7th European Conf. on Computer Vision*. London: Springer-Verlag, 2002. 67–81.
- [11] Huang C, Ai HZ, Wu B, Lao SH. Boosting nested cascade detector for multi-view face detection. In: *Proc. of the 17th Int'l Conf. on Pattern Recognition (ICPR 2004)*. IEEE Computer Society, 2004. 415–418.
- [12] Luo HT. Optimization design of cascaded classifiers. In: *Proc. of the Conf. Computer Vision and Pattern Recognition (CVPR 2005)*. Cambridge: IEEE Computer Society, 2005. 480–485.
- [13] Hou XW, Liu CL, Tan TN. Learning boosted asymmetric classifiers for object detection. In: *Proc. of the Conf. Computer Vision and Pattern Recognition (CVPR 2006)*. Cambridge: IEEE Computer Society, 2006. 330–338.
- [14] Brubaker SC, Mullin MD, Rehg JM. Towards optimal training of cascaded detectors. In: Leonardis A, Bischof H, Pinz A, eds. *Proc. of the 9th European Conf. on Computer Vision*. London: Springer-Verlag, 2006. 325–337.
- [15] Jin HL, Liu QS, Lu HQ, Tong XF. Face detection using improved LBP under Bayesian framework. In: *Proc. of the 3rd Int'l Conf. on Image and Graphics (ICIG 2004)*. 2004. 306–309.
- [16] Fang YC, Wang YH, Tan TN. Improving face detection through fusion of contour and region information. *Chinese Journal of Computers*, 2004,27(4):482–491 (in Chinese with English abstract).
- [17] Vapnik V. *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
- [18] Lu XG, Jain AK. Resampling for face recognition. In: Kittler J, Nixon MS, eds. *Proc. of the 4th Int'l Conf. on Audio- and Video-Based Biometric Person Authentication*. LNCS 2688, Heidelberg: Springer-Verlag, 2003. 869–877.
- [19] Zhou ZH, Jiang Y. NeC4.5: Neural ensemble based C4.5. *IEEE Trans. on Knowledge and Data Engineering*, 2004,16(6):770–773.
- [20] Chen J, Chen XL, Gao W. Face samples expanding based on the GA re-sampling. *Journal of Software*, 2005,16(11):1894–1901 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/16/1894.htm>
- [21] Wang RP, Chen J, Shan SG, Chen XL, Gao W. Enhancing training set for face detection. In: *Proc. of the Int'l Conf. on Pattern Recognition (ICPR 2006)*, Vol.3. Cambridge: IEEE Computer Society, 2006. 477–480.
- [22] Burges CJC. Simplified support vector decision rules. In: *Proc. of the 13th Int'l Conf. on Machine Learning*. 1996. 71–77.

- [23] Schölkopf B, Mika S, Burges CJC, Knirsch P, Müller KR, Rätsch G, Smola AJ. Input space vs. feature space in kernel-based methods. *IEEE Trans. on Neural Networks*, 1999,10(5):1000–1017.
- [24] Romdhani S, Torr P, Schölkopf B, Blake A. Computationally efficient face detection. In: *Proc. of the ICCV*. Cambridge: IEEE Computer Society, 2001. 695–700.
- [25] Chang CC, Lin CJ. LIBSVM: A library for support vector machines. 2001. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [26] Wang LW, Zhang Y, Feng JF. On the Euclidean distance of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2005, 27(8):1334–1339.
- [27] Simard P, Cun YL, Dender J. Efficient pattern recognition using a new transformation distance. In: Hanson S, Cowan J, Giles C, eds. *Proc. of the Advances in Neural Information Processing Systems*. San Mateo: Morgan Kaufman Publishers, 1993. 50–58.
- [28] Huttenlocher DP, Klanderman GA, Rucklidge WJ. Comparing images using the Hausdorff distance. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1993,15(9):850–863.
- [29] Wakahara T, Kimura Y, Tomono A. Affine-Invariant recognition of gray-scale characters using global Affine transformation correlation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2001,23(4):384–395.
- [30] <http://cbcl.mit.edu/software-datasets/FaceData2.html>

附中文参考文献:

- [2] 梁路宏,艾海舟,徐光祐,张钊.人脸检测研究综述. *计算机学报*,2002,25(5):449–458.
- [16] 方昱春,王蕴红,谭铁牛.融合人脸轮廓和区域信息改进人脸检测. *计算机学报*,2004,27(4):482–491.
- [20] 陈杰,陈熙霖,高文.基于遗传算法重采样的人脸样本扩张. *软件学报*,2005,16(11):1894–1901. <http://www.jos.org.cn/1000-9825/16/1894.htm>



王瑞平(1981—),男,山西孟县人,博士生,主要研究领域为计算机视觉,模式识别,机器学习.



陈熙霖(1965—),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为多模式人机接口,图像处理,模式识别,计算机视觉,数字电视技术.



陈杰(1975—),男,博士,主要研究领域为模式识别,计算机视觉,机器学习,数字水印.



高文(1956—),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为计算机视觉,多媒体,图像处理,多模式接口,人工智能,虚拟现实.



山世光(1975—),男,博士,副研究员,主要研究领域为图像处理与理解,模式分析与识别,机器学习,人机交互技术.